DOCTOR OF PHILOSOPHY

The genetic determinants of cardiovascular disease in patients with type 2 diabetes

Van Zuydam, Natalie

Award date:
2013

DOCTOR OF PHILOSOPHY

The genetic determinants of cardiovascular disease in patients with type 2 diabetes

Natalie Van Zuydam

2013

University of Dundee

The Genetic Determinants of Cardiovascular Disease in Patients with type 2 Diabetes

Natalie van Zuydam

For the degree of PhD

University of Dundee

January 2013

Table of Contents

List of Tables

List of figures

List of appendices

List of Abbreviations

Abbreviation	Meaning
25OHD	25-hydroxyvitamin D
ABI	Ankle brachial index
ADA	American diabetes association
ARIC	Atherosclerosis Risk in Communities
BMI	Body mass index
BNF	British National Formulary
CABG	Coronary artery bypass grafting
CAD	Coronary artery disease
CARDIoGRAM	Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis
CM	CardioMetabochip
CRP	C-reactive protein
CSF	Colony stimulating factors
CVD	Cardiovascular disease
CYP2R1	Cytochrome P450 2R1
DARTS	Diabetes Audit and Research in Tayside Scotland
DBP	Diastolic blood pressure
DHCR7	7-dehydrocholesterol reductase
eMERGE	Electronic MEdical Records and GEnomics centre
EMR	Electronic medical records
ESIST	The electronic stroke information system for Tayside
FGLU	fasting glucose
FHS	Framingham Heart Study
GC	Vitamin D binding protein
GLU	Fasting and non-fasting glucose
GoDARTS	Genetics of Diabetes and Audit Research in Tayside Scotland
GRO	General Register Office
GS	Genetic scores
GWAS	Genome wide association studies
HbA1c	Glycated haemoglobin
HDL-C	High-density lipoprotein cholesterol
HEARTS	Heart-disease Evidence-based Audit & Research Tayside Scotland
HIC	Health Informatics Centre
HMM	Hidden Markov model
HPC	high performance cluster
HWE	Hardy-Weinberg equilibrium
IBD	Identical by descent
ICD	International Classification of Diseases
IFN-y	Interferon gamma
IGF-1	Insulin like growth factor
IL-1	Interleukin 1

IL-2	Interleukin 2
IMI	Innovative medicines initiative
IS	Ischaemic stroke
ISGC	The international stroke genetics consortium
LDL-C	Low density lipoprotein cholesterol
LEAD	Lower extremity arterial disease
MCP-1	Monocyte chemo attractant protein 1
MGP	Matrix Gla protein
MI	Myocardial infarction
NO	nitrous oxide
NRI	Net reclassification improvement
oxLDL	Oxidised LDL-C
PDGF	Platelet derived growth factor
PSD	Practitioner Services Division
PTCA	Percutaneous coronary intervention
PTH	Parathyroid hormone
RCT	Randomised control trials
RE2	Random effects model
SBP	Systolic blood pressure
SCIDC	Scottish Care Information – Diabetes Collaboration
SHAPE-IT	Segmented haplotype estimation and imputation tool
SMC	Smooth muscle cells
SMR01	Scottish Morbidity Register
SUMMIT	SUrrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools (SUMMIT)
T1DM	Type 1 diabetes
T2D	Type 2 diabetes
TBI	Toe brachial pressure indexes
TCHOL	Total cholesterol
TG	Triglycerides
TGF-ß	Transforming growth factor beta
TNF-a	Tumour necrosis factor alpha
TZD	Thiazolidinedione
WHO	World health organisation
WTCCC2	Wellcome trust case control consortium 2

Acknowledgements

I acknowledge the support of the Health Informatics Centre, University of Dundee for managing and supplying the anonymised data and NHS Tayside, the original data owner. I am grateful to all the participants who took part in the Go-DARTS study, to the general practitioners, to the Scottish School of Primary Care for their help in recruiting the participants, and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. I would like to thank all the groups that submitted their own data which contributed to work presented in this thesis.

Part of this work was funded by charitable funding from Tenovus Scotland (grant number T10/36) and by IMI SUMMIT study, under the EU Framework Programme 7 funding stream. The Wellcome Trust provides support for Wellcome Trust United Kingdom Type 2 Diabetes Case Control Collection (Go-DARTS) and the Scottish Health Informatics Programme. Further informatics support is provided by the Chief Scientist Office of Scotland.

I would like to thank my supervisor, Prof Colin Palmer for his guidance and support throughout my PhD. I would also like to thank Dr Louise Donnelly, Dr Alex Doney and Prof Helen Colhoun for their academic and personal support during my PhD. I thank my mom, sisters, brother, Danial James, Rosie Nel and my Thursday/yoga dinner girls for their support through this process.

Declaration of the Candidate

I declare that I am the author of this thesis. All references have been consulted. The thesis is my own work, and has not been previously submitted for a higher degree.

Natalie van Zuydam

Declaration of the Supervisor

I certify that Natalie van Zuydam has completed the equivalent of nine terms of experimental research and that she has fulfilled the conditions of the University of Dundee so that she is qualified to submit this thesis in application for the degree of Doctor of Philosophy.

Professor Colin Palmer

Professor Andrew Morris

Work contributed by the candidate

All data files linked to the Genetics of Diabetes and Audit Research Tayside Scotland cohort were cleaned and manipulated from the raw data to derive clean data sets for statistical analysis. Similarly all the high density genotyping data were cleaned by the candidate and any manipulation including imputation of the genotyping data was conducted by the candidate. The high density genotyping chip experiments were conducted at external sites and the final data set was returned to GoDARTS so these were not generated by the candidate. However, all TaqMan genotyping was performed by the candidate as well as the downstream data cleaning.

For the purpose of statistical analysis all covariates including the genetic risk scores were derived by the candidate. All the statistical analysis was performed by the candidate. All the files contributed on behalf of GoDARTS to meta-analyses were also prepared by the candidate. In chapter 4, the summary statistics were prepared by the candidate and submitted for meta-analysis so the meta-analysis was not conducted by the candidate. The genetic score for coronary artery disease was derived by the candidate and its evaluation was also carried out by the candidate.

In chapter 5 the summary statistics from GoDARTS were prepared by the candidate but other files of summary statistics from contributing studies were prepared by individual analysts. All the subsequent file cleaning and combination of the files into a fixed and random effects meta-analysis were conducted by the candidate. Annotation of the meta-analysis associations were also performed by the candidate. These contributions are the same for Chapter 7. In chapter 6 the TaqMan genotyping, derivation of data sets and covariates for statistical analysis were all performed by the candidate.

Summary

The aim of this thesis was to identify genetic determinants associated with cardiovascular disease. We developed and verified algorithms to identify Coronary artery disease (CAD) , ischaemic stroke (IS) and lower extremity arterial disease (LEAD) cases and controls from electronic medical record data linked to Genetics of Diabetes and Audit Research Tayside Scotland cohort. We showed that these could be used in the discovery of novel genetic variants by replicating known signals in the *LPA* and the 9p21 region. We also identified unique associations of genetic scores for type 1 diabetes and triglycerides with the derived cardiovascular phenotypes. We used the derived CAD phenotype to contribute towards a large scale meta-analysis of CAD that led to the discovery of 15 new loci for CAD. A genetic score, that combined all the SNPs that passed a false discovery rate of 0.5% in the CAD meta-analysis, was not more predictive of CAD low density lipoprotein and high-density lipoprotein cholesterol. We found that a genetic score for decreasing 25-hydroxy vitamin D was associated with an increased risk of CAD. The effects of the genetic score suggested that there is a causal relationship between low 25OHD levels and increased CAD risk. A meta-analysis of CAD in patients with type 2 diabetes identified two independent signals in *ADAMTS7* associated with CAD at genome wide significance. Tests for heterogeneity of allelic effects for known CAD loci showed significant heterogeneity for signals in the 9p21 region, *ADAMTS7, ABO*, and *VEGFA.* A meta-analysis of LEAD replicated known associations in the 9p21 region and a known locus for nicotine dependence *CHRNA3.* A smoking interaction analysis identified signals in *ADAMTS17* that interact with smoking status to increase the risk of LEAD in non-smokers. We identified novel loci for CAD and putative loci for CAD in patients with T2D and LEAD with further analysis and replication required to establish these loci.

Chapter 1: Introduction

1.1 Cardiovascular disease

Cardiovascular disease (CVD) has many underlying aetiologies which include atherosclerosis, aortic stenosis, left ventricular hypertrophy and cardiomyopathy. The most common cause of CVD in the Western world is atherosclerosis of blood vessels in different parts of the vascular system. Cardiovascular diseases caused by atherosclerosis encompass coronary artery disease (CAD), ischaemic stroke (IS) and lower extremity arterial disease (LEAD). IS is caused by occlusion of blood vessels that supply the brain with blood; CAD in most cases results in myocardial infarction (MI) that occurs as the result of a coronary artery becoming blocked following the rupture of an atherosclerotic plaque[8] and LEAD is a chronic disease where the major vessels of the lower extremities become lined with atherosclerotic plaques and become progressively occluded[9].

Type 2 diabetes (T2D) increases the risk of CAD 2-3 fold in the general population[10] and is even higher in individuals who have T2D. T2D individuals are at much greater risk of CAD, LEAD and IS[10, 11] where CVD complications account for half the deaths in patients with T2D[12]. It has been reported that the incidence of atherosclerosis is 3-4 fold higher in the T2D population compared to non-diabetic individuals[13], which is largely attributed to insulin resistance. Hyperinsulinemia causes natural metabolic processes to become imbalanced which has a knock on effect in other pathways. Circulating levels of glucose and triglycerides increase which leads in time to dyslipidaemia[14, 15]. This places a large amount of stress on the endothelium and eventually the endothelium becomes dysfunctional and the atherosclerotic process begins[14, 15].

1.2 Atherosclerosis

Atherosclerosis is a progressive multifactorial disease characterised by the accumulation of lipids and fibrous elements in the large arteries[16, 17]. Early atherosclerosis is characterised by the development of atherosclerotic lesions and four main processes are involved their formation: accumulation of lipids, free cholesterol and esterified cholesterol in the matrix of the intima; proliferation and chemo-attraction of smooth muscle cells (SMC), macrophages which leads to the formation of

foam cells; the release of inflammatory markers and synthesis of a connective tissue matrix by SMCs[18, 19]; and finally SMC proliferation to form a fibrous cap.

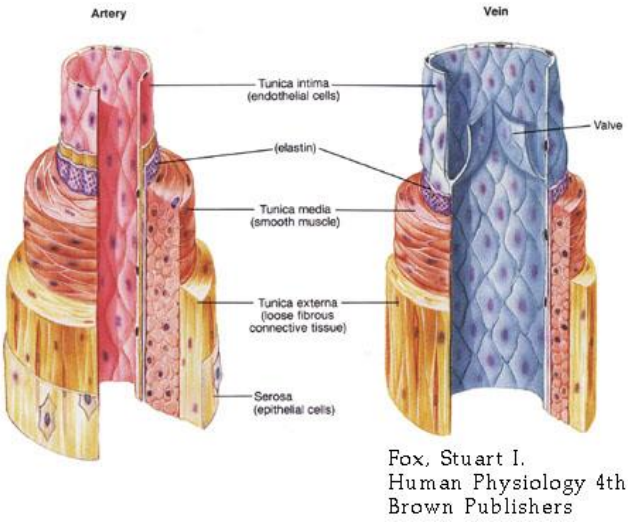

Figure 1-1: Structure of artery and vein walls where the endothelium is the barrier between blood and tissue.

1.2.1 Accumulation of lipid particles

The first step in atherosclerosis is the accumulation of lipid particles in the media (Figure 1-1). Two main theories have been proposed to explain the initiating step of atherosclerosis: the "response-to-injury" hypothesis proposed by Ross et al., 1977 and the "response-to-retention" hypothesis proposed by Williams et al., 1995. Both theories indicate that forces acting on the endothelium (Figure 1-1) are responsible for the initiation of atherosclerosis.

1.2.1.1 Response-to-injury

The Response-to-injury hypothesis proposes that the development of atherosclerotic lesions is a result of injury to the endothelium. Injury to the endothelium may be caused by shear stress caused by the flow of blood through the vessel or may be caused by biochemical agents in the blood such as lipids, Homocysteine and uraemia. Infections and immunologic injury have also play a role in endothelial injury[20]. Injury to the endothelium causes changes in the morphology of the endothelial cells, which prevents the cells from forming a continuous layer. Injury may result in the loss of endothelial cells from the endothelial layer and subsequent exposure of the underlying collagen that is then bound by platelets.

The endothelium becomes permeable to platelet derived factors and lipoproteins and these infiltrate the endothelium and migrate through to the media. This invasion of particles causes the vascular SMCs to migrate into the media and to proliferate. The SMCs also produce new connective tissue which increases the deposition of extra cellular lipids causing the lesion to enlarge. Repeated injury leads to larger atherosclerotic plaques that may hinder blood flow[20].

1.2.1.2 Response-to-Retention

The "response-to-injury" theory has been superseded by the "response-to-retention" hypothesis which proposes the retention of lipids as the first step of atherosclerosis. This theory is a plausible explanation as atherosclerotic plaques are often covered by an intact endothelial layer[21].

Transport of low density lipoprotein cholesterol (LDL-C) into and out of the media takes place naturally across a healthy endothelium. Some of the LDL-C is retained in the media while some is released back into the blood stream. When there are normal serum levels of LDL-C in the blood stream the rate of LDL-C influx does not exceed the rate of efflux and only small amounts of LDL-C are retained. When serum levels of LDL-C are high more LDL-C particles are retained in the extra-cellular matrix than are released back into the blood stream. The LDL-C particles are bound to proteoglycans[22] by apoB which is found in both LDL-C and lipoprotein (a) [21]. LDL-C retention is also increased by lipoprotein lipase which is produced by the endothelial cells increases the uptake of lipoproteins from the blood stream.

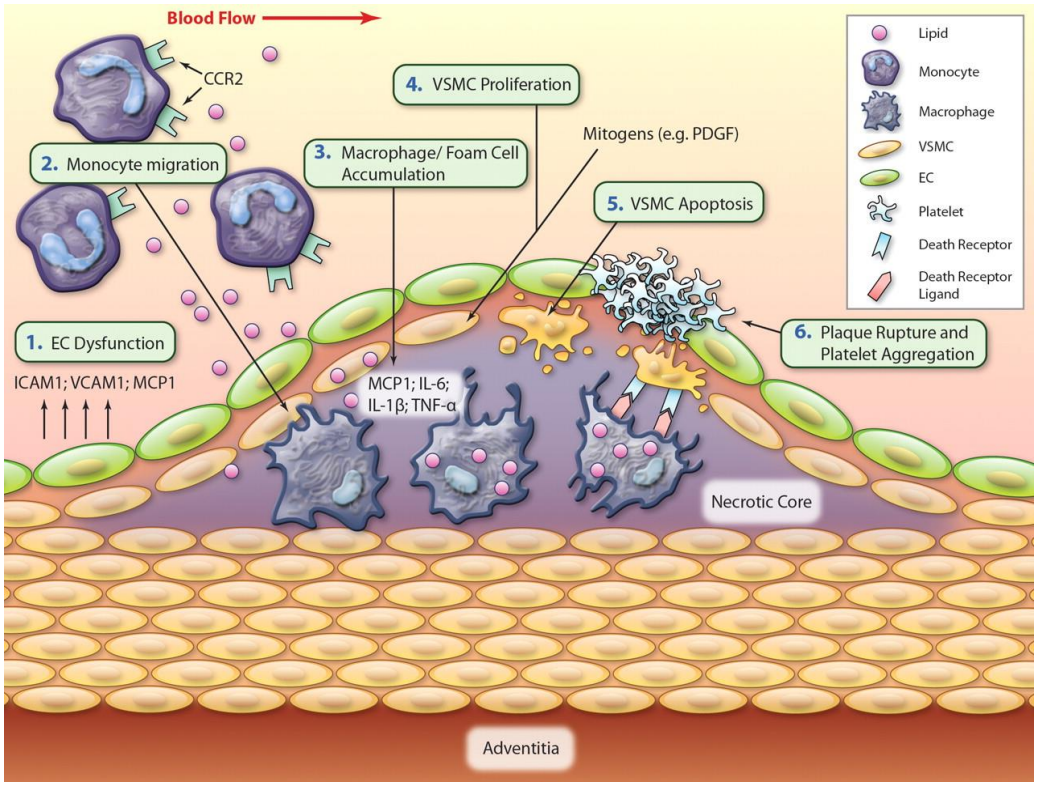

Figure 1-2: Schematic of atherogenesis and an unstable atherosclerotic plaque[23]

1.2.2 Foam cell formation

Once lipids are bound to the extra cellular matrix either through retention or as a response to injury it stimulates a number of biological pathways to produce an atherosclerotic lesion. Retained LDL-C is oxidised by oxidising species that are naturally produced by healthy cells surrounding the media. The oxidation of LDL-C causes the endothelial cells and the vascular SMCs to express monocyte, T-lymphocyte and SMC chemoattractants[21] such as colony stimulating factors (CSF), monocyte chemo attractant protein 1 (MCP-1) and transforming growth factor beta (TGF-β).

The oxidised LDL-C (oxLDL) is taken up by macrophages and by vascular SMCs (Figure 1-2). The oxLDL is mainly bound by two scavenger receptors, SR-A and CD36, expressed on the surface of macrophages. The oxLDL is ingested by the macrophages and they become foam cells. The foam cells stimulate proliferation, migration of vascular SMCs and release of proteoglycans and lipoprotein lipase from the SMCs[21]. This increases the amount of LDL-C that is retained in the media. As the amount of retained LDL-C increases the diseased area expands into an atherosclerotic lesion[22].

1.2.3 Inflammation

LDL oxidation stimulates the endothelium to produce a number of pro-inflammatory markers like interleukin 1(IL-1), tumour necrosis factor alpha (TNF-α), interferon gamma (IFN-γ) and interleukin 2 (IL-2) with the CSF as well as adhesion molecules and growth factors[24]. The pro-inflammatory markers attract monocytes and leukocytes to the lesion (Figure 1-2). Adhesion molecules expressed by the endothelial cells allow the monocytes to bind the endothelium and invade the media[21]. As well as initiating the release of pro-inflammatory markers, oxidised LDL can also inhibit the production of nitric oxide by the endothelial cells that causes endothelial dysfunction.

1.2.4 Endothelial dysfunction

The endothelium plays an important role in maintaining vascular health as it function as a selective barrier to the diffusion of macromolecules from the lumen to the interstitial space; and through the production of nitrous oxide (NO), regulates vascular tone; modulates inflammation; promotes and inhibits of vascular growth and modulates platelet aggregation[25]. Any process which causes endothelial dysfunction has a large impact on the health of the blood vessel and plays a role in the development or acceleration of atherosclerosis. Oxidised LDL-C and the production of reactive oxygen species contribute to endothelial dysfunction.

NO has many anti-atherosclerotic properties and acts on the endothelium and SMCs to prevent atherosclerosis. NO interacts with prostacyclin to inhibit platelet aggregation and inhibits the adhesion of neutrophils to the endothelium. NO in high concentrations also inhibits the proliferation of smooth muscle cells. So under conditions of endothelial dysfunction atherosclerotic processes are accelerated (Figure 1-2). Inflammatory cells such as macrophages, monocytes, neutrophils and lymphocytes can infiltrate into the media and the proliferation of SMCs is not regulated[25].

1.2.4.1Measuring endothelial dysfunction

Endothelial dysfunction can be clinically assessed in coronary or peripheral arteries and or arterioles via intracoronary or intrabrachial infusion of the endothelium with acetylcholine. Non-invasive techniques can also be used such as flow mediated dilation of the brachial artery, finger-pulse plethysmography or by pulse curve analysis[25].

1.2.5 Fibrous cap formation

The final stages of atherosclerotic plaque maturation is characterised by the formation of a thin fibrous cap that covers the atherosclerotic plaque. Fibrous plaques are characterised by a growing mass of extracellular lipid, by the accumulation of SMC's and SMC derived extracellular matrix. SMCs are stimulated to migrate into the atherosclerotic lesion by platelet derived growth factor (PDGF) and insulin like growth factor (IGF-1). The migration and proliferation of SMCs causes the fatty streak to expand into a plaque. The plaque is stabilised by a fibrous cap made up of SMCs[24] (Figure 1-2).

The expanding plaque convolutes the luminal surface of the artery[26, 27] and hinders blood flow. During the advanced stages of atherogenesis the lesions become vascularised. Fissuring, cracking or ulceration of the plaque can lead to a haemorrhage from the lumen or from small vessels[26], which may manifest as an MI or an IS.

1.2.6 Atherosclerosis and type 2 diabetes

Atherosclerosis is accelerated in patients with insulin resistance, metabolic syndrome and T2D[28, 29]. The main causes are A-FLIGHT toxicity (Table 1-1) which cause injury to the vascular wall or where atherosclerosis is already present accelerate plaque formation. Individuals who suffer from insulin resistance, metabolic syndrome and T2D have increased levels of circulating reactive oxygen species which decreases the half-life of nitric oxide [25]. Patients with T2D have higher levels of circulating lipids which leads to increased retention of these particles in the media. They also have higher level of advanced glycation end products and may also suffer from hypertension increasing the chances of injury to the endothelium. which can cause the oxidation of lipids[16]. These factors combine with already present chronic inflammation to increase the amount of atherosclerosis present in the blood vessels[28, 29].

Table 1-1: The manifold toxicities of insulin resistance, metabolic syndrome and type 2 diabetes mellitus[28, 29].

	A-FLIGHT toxicities
A	Amylin (hyperamylinemia)/amyloid toxicity
	Ang II (also induces PKC)
	AGEs/AFEs (advanced glycosylation/fructosylation endproducts)
	Antioxidant reserve compromised
	Absence of antioxidant network
	Ageing
	Angiogenesis (induced redox stress) Arteriogenesis (impaired PAI-1)
	Atherosclerosis – Atherosclerosis. [ROS beget ROS]
F	Free fatty acid toxicity
L	Lipotoxicity
I	Insulin toxicity (hyperinsulinemia-hyperproinsulinemia) (endogenous)
	Inflammation toxicity
G	Glucotoxicity (compounds peripheral insulin resistance) reductive stress Sorbitol / polyol pathway
	Pseudohypoxia (NADH/NAD increased) PKC
H	Hypertension toxicity
	t homocysteine toxicity
T	Triglyceride toxicity

1.3 Genetics of Cardiovascular disease

1.3.1 Coronary artery disease

CAD occurs when the coronary arteries become narrowed or blocked by atherosclerotic plaques causing a lack of blood flow to the heart. This process can manifest itself as clinical symptoms in the form of shortness of breath, fatigue and general weakness as the heart is not pumping enough blood around the body. Narrowing of the blood vessels can cause angina pectoris that manifests as chest pain; blockage of the coronary arteries can result in an MI. An MI is irreversible necrosis of the heart muscle where lack of blood flow deprives part of the heart muscle of nutrients and oxygen resulting in damage and apoptosis of the heart muscle cells. Stenosis of the coronary arteries can be treated with different drugs or by interventions in the form of percutaneous coronary intervention (PTCA), where a stent is inserted at the site of the stenosis. The stenosis can also be bypassed by coronary artery bypass grafting (CABG) where blood vessels are harvested from other parts of the body and grafted onto the affected artery to bypass the blocked section of the coronary artery.

CAD is the leading cause of death worldwide and is the main macrovascular complication of T2D[30]. According to the British Heart Foundation, coronary artery

disease is the cause of premature death of 1 in 5 men and 1 in 7 women in the United Kingdom with the highest death rate in Scotland. The American Heart Association reports similar results in the United States of America where it is the main cause of death for 1/3 of deaths that occur. T2D is thought to affect 8.5% of the world population and more than 50% of individuals suffering from diabetes will succumb to a vascular complication of which the majority are cardiovascular. The heritable component of CAD has been estimated at between 40 and 50 % in twin and family studies[31] but large-scale genetic studies have only explained a small amount of the heritability.

1.3.2 The genetics of coronary artery disease

Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis[32] [32] is a consortium that has been formed to investigate the genetic determinants of CAD and myocardial infarction[33]. CARDIoGRAM incorporates published and unpublished genome wide association studies (GWAS) studies from Atherosclerotic Disease VAscular functioN and genetiC Epidemiology study[34-37], CADomics[38], Cohorts for Heart and Aging Research in Genomic Epidemiology[39], deCODE [40], the German Myocardial Infarction Family Studies I, II, and III [41], Ludwigshafen Risk and Cardiovascular Health Study/AtheroRemo[42], MedStar[43], Myocardial Infarction Genetics Consortium[44], Ottawa Heart Genomics Study[45], PennCath[46], and the Wellcome Trust Case Control Consortium[47] amounting to data on over 22000 CAD cases and 60000 CAD free controls[33].

There are 25 loci that have been reported at genome wide significance for CAD by Schunkert et al., 2011 for CARDIoGRAM and another six by other consortia[38, 44, 48-52]. 12 of the loci had been published in previous studies by members of CARDIoGRAM, 13 loci were identified by combining the data across CARDIoGRAM (Figure 1-3) and 6 by other consortia which include the IBC CAD consortium. The 25 loci identified in the CARDIoGRAM study[48] explain about 10% of the total heritability of CAD[48]. The CAD associated SNPs are found in genes that also contain signals for associated risk factors such lipid metabolism, blood pressure, glucose homeostasis, chemotaxis and in other genes that are involved in regulating cell growth and cycle(Table 1-2). The genetic determinants have clear roles in atherosclerosis: OxLDL is the main component of atherosclerotic plaques[26] as many of the genes associated with CAD also play a role in

the metabolism of LDL-C and other signals are found in genes that affect cell adhesion, migration, cell death and cellular matrix formation may have roles in the formation and stability of atherosclerotic plaques (Table 1-2).

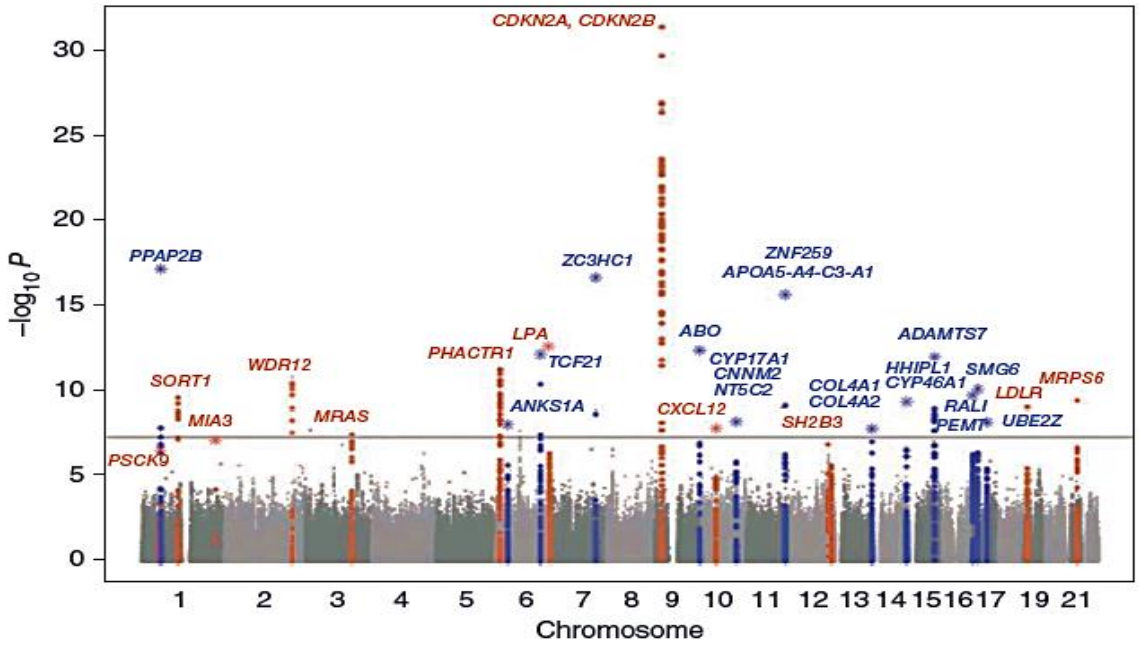

Figure 1-3: A Manhattan plot taken from Schunkert et al., 2011 annotated for previously described CAD loci (red) and those that were reported as new loci (blue). Data from the discovery phase are shown in circles, and data from the combined discovery and replication phases are shown in stars[48].

Table 1-2: Thirty one loci have been identified by the CARDIoGRAM consortium, the IBC CAD consortium and in other smaller studies have a number of different biological functions and the genes contained signals for other phenotypes[32, 38, 44, 48, 50-52].

Locus	SNP	Closest Gene	Biological function	Other phenotypes
1p32.3	rs11206510	*PCSK9*	Lipid homeostasis	LDL- C levels[53]
1p13.3	rs599839	*SORT1*	Lipid homeostasis	LDL-C and TC[54]
1q41	rs17465637	*MIA3*		
2q33.1	rs6725887	*WDR12*	Ribosome biogenesis and cell proliferation[55]	
3q22.3	rs2306374	*MRAS*	Cell growth and differentiation[56]	
6p24.1	rs12526453	*PHACTR1*	Endothelial cell death and tube formation[57]	
6q25.3	rs3798220	*LPA*	Lipoprotein A levels[58]	
7q22	s10953541	*Gene desert*		
9p21.3	rs4977574	*CDKN2A, CDKN2B*		Glioma[59]
10q11.21	rs1746048	*CXCL12*	Lymphocyte activation	Cancer[60]
12q24.12	rs3184504	*SH2B3*	Cell signalling[61]	Type 1 diabetes[62]; Blood pressure[63]
19p13.2	rs1122608	*LDLR*	Lipid homeostasis	LDL-C levels[64]
21q22.11	rs9982601	*MRPS6*	Mitochondrion ribosome	Respiratory function[65]

Locus	SNP	Closest Gene	Biological function	Other phenotypes
			protein	
1p32.2	rs17114036	*PPAP2B*	Cell migration and adhesion[66]	
6p21.31	rs17609940	*ANKS1A*		
6q23.2	rs12190287	*TCF21*	Transcription factor	Lung cancer[67]
7q32.2	rs11556924	*ZC3HC1*		
9q34.2	rs579459	*ABO*		E-selectin[68]; Depression[69]; Erythrocyte indices[70]
10q24.32	rs12413409	*CYP17A1,CNNM2, NT5C2*	Epithelial transport, purine metabolism.	Blood pressure[71]; Schizophrenia
11q23.3	rs964184	*ZNF259, APOA5-A4-C3-A1*	Lipid homeostasis[54]	Metabolic syndrome[73]; Triglycerides[64]
13q34	rs4773144	*COL4A1, COL4A2*	Stability of basement membranes[74]	Cerebral small vessel disease[74]
14q32.2	rs2895811	*HHIPL1*		
15q25.1	rs3825807	*ADAMTS7*	Degrades cartilage matrix protein[75]	
15q25.1	Rs4380028	*ADAMTS7*		
17p13.3	rs216172	*SMG6, SRR*	Telomerase stability	Aorta[76]; Type 2 diabetes[77]
17p11.2	rs12936587	*RASD1, SMCR3,PEMT*	Cell morphology and growth	
17q21.32	rs46522	*UBE2Z, GIP, ATP5G1, SNF8*	Glucose homeostasis	
10q23.2	rs1412444	*LIPA*	Lipid homeostasis	
11q22.3	rs974819	*PDGFD*	Cell migration[78]	

Locus	SNP	Closest Gene	Biological function	Other phenotypes
10p11.23	rs2506083	*KIAA1462*	Cell adhesion[79]	
19q13.2	rs2075650	*ApoE-ApoC1*	Lipid homeostasis	C-reactive protein[80]

Large studies need to be conducted to identify variants that explain the missing heritability of CAD. Although diabetes is a large predictor of CAD risk[13] the two diseases do not share any common genetic determinants[48]. Analyses to identify genes that specifically determine CAD in T2D individuals and in non-diabetic individuals separately may provide insights into the role of T2D in CAD.

1.3.3 Ischaemic stroke

Ischaemic stroke is a broad description of a number of stroke subtypes that have different underlying aetiologies. Blockages in the blood supply to the brain cause a rapid loss of brain function manifesting in clinical symptoms that include a loss of eyesight, motor function and speech. Acute IS can be divided into 5 subtypes: large artery atherosclerosis, cardioembolism, small-vessel occlusion, stroke of other determined aetiology and stroke of undetermined aetiology. Individuals who suffer a large-artery stroke have stenosis (>50%) or occlusion of the main brain artery and usually have a history of atherosclerotic disease[81]. Cardioembolic stroke is due to arterial occlusions caused by an embolus arising in the heart. Individuals who suffer from atrial fibrillation, sick sinus syndrome, myocardial infarction (less than 4 weeks), and left ventricular thrombus are at higher risk of cardioembolic stroke[82]. Small-vessel stroke occurs due to occlusions of the small arteries of the brain. Individuals who suffer from small-vessel stroke are usually hypertensive and have diabetes mellitus[82]. Strokes with other determined aetiology do not fit into the above categories but have an identifiable reason for the stroke such as haematological disorders and hypercoagulable states[82]. Stroke of undetermined aetiology have no known cause of the stroke[83].

1.3.4 The genetics of ischaemic stroke

The international stroke genetics consortium (ISGC, http://www.strokegenetics.org/) is a consortium similar to CARDIoGRAM which combines data from published and

unpublished GWAS on stroke. Since the stroke subtypes have such distinct aetiologies large genetic studies on ischaemic stroke have been unsuccessful in identifying common genetic determinants of ischaemic stroke. Studies that have been targeted at identifying variants associated with stroke subtypes have identified the 9p21 region associates with large artery stroke[84] and variants on 4q22 and 16q22 with cardioembolic stroke[85, 86]. Atherosclerosis of the carotid arteries has also been associated with the 9p21 region and is the underlying cause of large artery stroke. *PITX* is known to cause atrial fibrillation[87] the underlying cause of cardioembolic stroke , as is *ZFHX3*[88] (Table 1-3).

Candidate gene studies have identified loci that correlate with risk factors for IS. *APOE, LPA* and *eNOS* are associated with atherosclerotic disease[48, 89, 90], the major underlying cause of IS[91, 92]. Other loci that influence blood pressure like *ACE* and *MTHFR* are likely candidates as one of the predictors of IS is hypertension[92]. The candidate gene loci also include loci that could be linked to drug metabolism like the cytochrome p loci (Table 1-3). *VCKOR1*A explains a large proportion of the variation of Warfarin dose an anticoagulant commonly prescribed to treat IS[93] (Table 1-3).

Table 1-3: Candidate gene loci and genome wide association study loci for ischaemic stroke

Study	Locus	Gene	SNP	Biological function	Other phenotypes
Candidate genes	17q23.3	*ACE*	[94]	Blood pressure homeostasis	Response to *ACE* inhibitors[95]
	13q12	*ALOX5AP*	rs17216473	Inflammatory Response	Asthma[96]
	8q23.1	*ANGPT1*	rs2507800[97, 98]	Angiogenesis	
	6q26	*APOA* [99]	[75, 76]	Lipoprotein A	Coronary artery disease[48]
	19q13.2	*APOE*	[100]	Triglyceride Metabolism	C-reactive protein[80] and atherosclerosis[90]
	1q21-23	*CRP*	rs2794521[101, 102]	Host defence	C-reactive protein[103]
	1p33	*CYP4AII*	[104]	Drug and lipid metabolism	
	19q13.12	*CYP4F2*	rs2108622[104]	Drug and lipid metabolism	Vitamin E[105]
	8q21-q22	*CYPIIB2*	rs1799998[106, 107]	Drug and lipid metabolism	
	1p22	*DDAH1*	[108]	Nitric oxide Metabolism	
	7q36	*eNOS*	Rs1799983[109]	Vasodilation	Atrial fibrillation[110] and coronary artery disease[89]
	1q23	*F5 LEIDEN*	[94]	Blood coagulation	Venous thrombosis[111], D-Dimer[112] and P-selectin[69]
	4q28	*FGB*	[113]	Blood coagulation	Fibrinogen[114, 115]
	17p3-12	*GP1BA*	[116]	Blood Coagulation	Platelet count[70]
	17q21.32	*ITGB3*	[117]	Cell adhesion and signalling	Glanzmann thrombasthenia[118]
	7p21	*IL6*	[119, 120]	Inflammation	
	5q35	*LTC4S*	rs730012[121, 122]	Production of leukotrienes	

Study	Locus	Gene	SNP	Biological function	Other phenotypes
	1p36.3	*MTHFR*	rs2274976[94]	Homocysteine methylation	Homocysteine[123] and blood pressure[63]
	7p15.1	*NPY*	rs16147[124]	Cardiovascular function	
	7q22.1	*PAI-1*	[125]		
	7q21.3	*PAROXONASE -1*	rs662[126]	Lipid homeostasis	Prevents oxidation of HDL in the liver
	1p32.3	*PCSK9*	rs11206510[127]	Lipid homeostasis	Low density lipoprotein cholesterol[53]
	5q12	*PDE4D*	rs12188950[128, 129]	Signal transduction	Asthma[130], sleep[131] and oesophageal carcinoma[132]
	11p11	*PROTHROMBIN*	[94]	Blood Coagulation	
	6q23	*SGK1*	rs1057293[133]	Stress response	
	6p21.3	*TNF-α*	[134, 135]		Type 2 diabetes[136]
	16p11.2	*VKORC1*	rs9923231[137]	Blood Coagulation	Warfarin[93]
GWAS		*9p21*	rs4977574[126]	Cell cycle	Coronary artery disease[40]
	12p13	*NINJ2*	rs11833579[138, 139]	Cell adhesion	
	14q23.1	*PRKCH*	rs2230500[140, 141]	Cell signalling	
	4q25	*PITX2*	rs2200733[142]	Transcription factors	cardioembolic stroke[143]
	16q22.3	*ZFHX3*	rs7193343[144]	Cell cycle	Atrial fibrillation[144]; cardioembolic stroke[143]
	7p21.1	*HDAC9*	rs11984041[145]	Chromosome structure	Body weight[146] and electrocardiography[147]

Candidate gene studies have only included a few hundred IS cases and are comparatively smaller than large genome wide association studies. Thus these candidate genes need to be confirmed in larger genetic studies. Currently, only 5 loci have been identified by genome wide association studies (Table 1-3). This may indicate that larger sample sizes are required to identify more variants. It is pertinent to

consider that IS is made up of 5 subtypes three of which have well defined and distinct aetiologies so GWAS may be more successful if they were focused on the subtypes rather than the broader IS phenotype.

1.3.5 Lower extremity arterial disease

It has been estimated that lower extremity arterial disease affects between 4.5 and 12% of the population[148,149-151]. Prevalence increases with age where it is up to 20% in people over 70. The underlying atherosclerotic disease increases the chances of mortality from other cardiovascular diseases by six fold[152] making it an important indicator of CVD death. Coronary artery disease is the main cause of death in individuals with LEAD followed by cerebrovascular disease and to a lesser extent by other cardiovascular disease such as ruptured aneurysms[153-155]. The disease is characterised by atherosclerotic occlusions of the larger arteries of the lower limbs.

The most common method of diagnosis is an abnormal ankle brachial index (ABI) measurement of less than 0.9 or greater than 1.3. A low ABI usually indicates limb ischaemia while an abnormally high ABI is characteristic of vessel stiffening by calcification of the media[156]. Partial stenosis of the large arteries results in intermittent claudication and may be exacerbated by exercise. In extreme cases complete stenosis of the large arteries completely blocks blood flow, sores and ulcers develop and eventually gangrene[156]. Gangrenous infections are treated by amputating the affected limb, while intermittent claudication can be treated using drugs like pentoxifyline and cilostazol[156]. Vascular intervention surgeries can also be used to remove or bypass blocked vessels[156, 157].

1.3.6 The genetics of lower extremity arterial disease

The heritability of ABI has been estimated from family studies to be between 21 and 30%[158]. Currently there are only two genome wide association studies that have been published for LEAD[159, 160]. Two signals were identified at genome wide significance: in *CHRNA3,* which is an established locus for nicotine dependence and in the well-established atherosclerosis locus 9p21 which is associated with CAD and IS[48, 160, 161]. Rs3798220 and rs10455872 in the *LPA* have also been associated with LEAD in the Heart Protection Study[162].

There are 19 loci which were reported in candidate gene studies; these are listed in Table 1-4. The SNPs are found in genes which are linked to blood pressure, blood coagulation, cell signalling, lipid metabolism, calcification, inflammation and heart disease. *ACE, AGT* and *MTHFR* have been associated with blood pressure and hypertension[63, 95, 163], known risk factors for LEAD[164]. Vessel calcification is well characterised for LEAD so signals in *ENPP1,* which has been associated with plaque and vessel calcification[165, 166], are also good candidates. *CSMD1* is an established CAD locus a risk factor for LEAD[164]. These associations have not been replicated in large populations so their association with LEAD still needs to be confirmed.

An international consortium of all the GWAS available for LEAD has not been established and the meta-analysis published by Murabito et al., 2011, which includes 3409 LEAD cases from 21 population based cohorts, is the largest collection of GWAS data for LEAD to date[160].

Table 1-4: Nineteen candidate SNPs have been associated with lower extremity arterial disease.

SNP	Gene/ Mutation	Biological function	Other phenotypes
rs4340	*ACE* indel[167]	Blood pressure homeostasis	Response to *ACE* inhibitors[95]
rs5051 /rs699	*AGT* [168] C/T[169]	Blood pressure Homeostasis	Hypertension[163]
rs2554503	*CSMD1*[170]	Carcinogenesis [171]	Coronary artery disease[172]
rs1044498	*ENPP1* K121Q[173]	Calcification	Plaque and vessel calcification[165] [166]
rs1801020	*F12*[174]	Blood coagulation	
rs6025	*F5* 1691 G/A[169]	Blood coagulation	Venous thrombosis[111], D-Dimer[112] and P-selectin[69]
rs1800790	*FGB* -455G/A[175]	Blood coagulation	Fibrinogen[114, 115]
rs1800795	*IL6* (174 G/C)[169]	Inflammation	
rs5918	*ITGB3*[169]	Cell adhesion and signalling	Glanzmann thrombasthenia[118]
rs3918242	*MMP9* (-1562 C/T)[169]	Extracellular matrix Metabolism	
rs1801133	*MTHFR* C677T[176]	Homocysteine methylation	Homocysteine[123] and blood pressure[63]
rs1800591	*MTTP*[177]	Lipid metabolism	Non-alcoholic fatty liver disease[178]
rs2070744	*NOS3*[179]	Vasodilation	Atrial fibrillation[110] and coronary artery disease[89]
rs891512	*NOS3*[180]		
rs1902341	*OSBPL10*[170]	Lipid receptor	
rs11591147	*PCSK9* R46L[181]	Lipid homeostasis	Low density lipoprotein cholesterol[53]
rs1805182	Pro12Ala *PPARG*[182]	Transcription	Obesity[183], diabetes[184] and atherosclerosis[185]
rs1799963	*PT* G20210A[186]	Blood coagulation	
rs235243	*VPS13D*[170]	Vacuolar-protein-sorting	

1.4 Risk factors for cardiovascular disease

CAD, IS and LEAD share several risk factors which include elevated blood pressure, diabetes, fasting glucose, total cholesterol, low density lipoprotein cholesterol, high-density lipoprotein cholesterol (HDL-C), triglycerides and smoking[164, 187-189]. These risk factors are well established in epidemiology and recently variants that predict cardiovascular risk factors have been identified in large genetic association studies.

Several of the studies have published genetic risk scores for traits and associated with outcomes such as CAD. These studies provide additional evidence that may be used to predict disease outcomes from genetic data.

1.4.1 Genetics of the risk factors for cardiovascular diseases

Ehret et al., 2011 identified 16 novel loci associated with systolic (SBP) and diastolic blood pressure (DBP) in addition to the 13 established SNPs[31]. The authors calculated a composite risk score for increasing SBP and DBP and found that it was associated with increased left ventricular wall thickness and increased risk of IS and CAD[31]. While CAD and T2D do not share any genetic signals, fasting glucose is correlated with cardiovascular risk[190], and both type 1 and type 2 diabetes are established cardiovascular risk factors.

GWAS for diabetes and glycaemic related traits have identified loci that are associated with type 1 diabetes, type 2 diabetes and fasting glucose. Dupuis et al., 2010 identified 17 loci that had a replicated allelic effect on fasting glucose. There are 42 loci that have been identified for type 1 diabetes which include the *HLA* locus identified in linkage studies, 5 genes from candidate gene studies and 36 signals identified by GWAS[191]. There are 48 established loci for T2D[31, 77, 192, 193] which have been identified in multi-ethnic populations. The relationship between genetic risk scores calculated from these variants and cardiovascular outcomes has not been established.

Lipid levels are the largest predictor of cardiovascular disease and play an important role in how cardiovascular disease is treated. 95 loci have been reported for lipid traits, 59 novel loci that were identified by Teslovich et al., (2010) and 36 previously established loci. Of the novel loci: 39 were associated with total cholesterol; 22 with LDL-C; 31 with HDL-C and 16 with triglycerides. The majority of previously identified loci associated with more than one lipid trait[54]. Teslovich et al, found that 14 of the 95 loci were associated with CAD after correcting for multiple testing, 12 of the SNPs had associations with other lipid traits and 4 of the lipid effect alleles had an opposite effect on CAD risk[54]. The findings of Teslovich et al 2010 indicate that despite an increase or decrease in a corresponding lipid trait the allelic effect does not always translate to increased risk of CAD. These findings need to be investigated further.

Smoking status is the biggest predictor of LEAD and is associated to a lesser extent with IS and CAD. Several loci have been associated with smoking behaviour in large genetic association studies. The Tobacco and Genetics consortium identified the following loci which were associated with the number of cigarettes smoked per day: *CHRNA3, LOC100188947, CYP2A6/EGLN3, BDNF* and *DBH*[194]. Thorgeirsson et al identified *CYP2A6/RAB4D, CHRNB3/CHRNA6,* 7p14.3 and *CYP2B6*[195]. *CHRNA3* is an established locus for LEAD but the effect of the other smoking loci on LEAD has not been assessed[159].

1.5 Vitamin D deficiency and cardiovascular disease

In addition to the conventional risk factors many other traits have been indicated as risk factors for CVD. Vitamin D deficiency has been identified as a risk factor for CAD and supplementation trials show that there is a decrease in overall mortality with vitamin D supplementation but this is not specific to CVD outcomes[6, 196, 197]. The causal relationship between vitamin D levels and CAD outcomes is unclear as vitamin D levels can be affected by many factors such as poor diet and lack of sunlight[196], therefore low vitamin D levels could be due to the illness rather than a cause.

1.5.1 Serum vitamin D concentrations decrease with increasing latitude

Healthy concentrations of serum vitamin D are accepted as between 70 and 250nmol/L[198], insufficiency at between 25-50nmol/L, deficiency at less than 25nmol/L and toxicity at greater than 250nmol/L. Insufficiency and deficiency have serious health consequences including myopathy, reduced bone mineral density, rickets and hyperparathyroidism[199-201]. Vitamin D insufficiency and deficiency increases with increasing distance from the equator[198] (Figure 1-4) where individuals from countries 40° North or South of the equator have vitamin D insufficiency or deficiency (Figure 1-4).

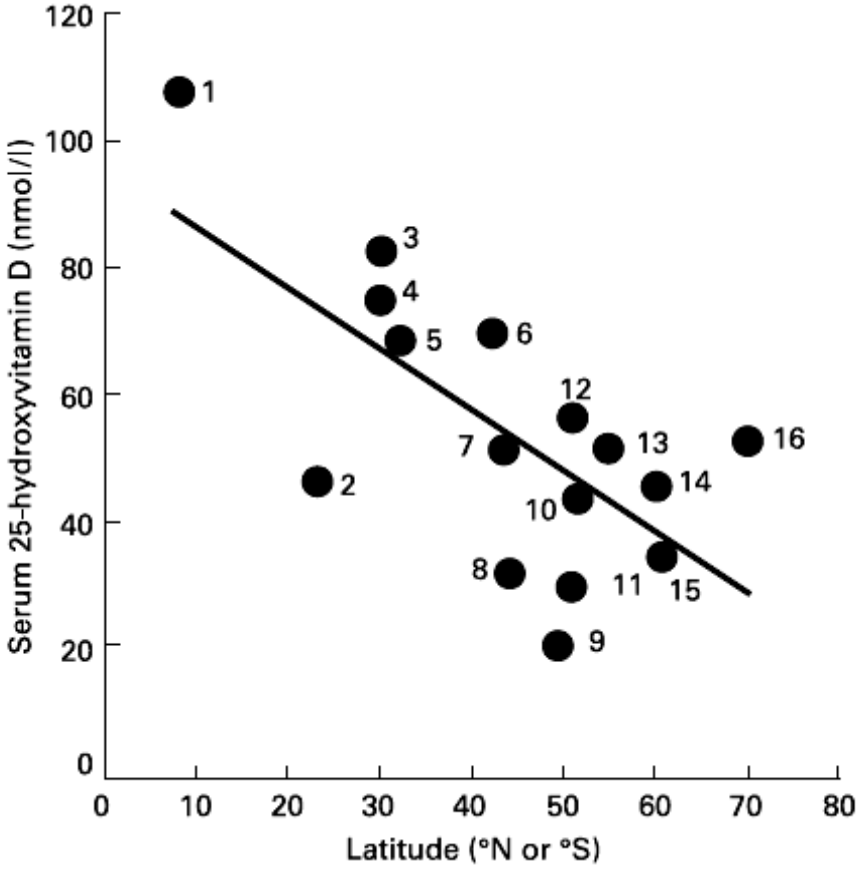

Figure 1-4: Mean circulating 25-hydroxyvitamin D levels in children, adolescents, and adults according to geographic latitude (r^2 0·68; P, 0·01). [191] Children; (2) male adults; (3) male adolescents and adults; (4) female adolescents and adults; [52] male adults; [168] adolescents and adults; (7) adults; (8) children; (9) adolescents; (10) children; (11) adults; (12) adults (13) children; (14) adults; (15) adolescents; (16) adults[198].

Interestingly, the same trend is true for increasing death rates from ischaemic heart disease (IHD) where the death rate increases with increasing latitude (Figure 1-5)[198]. Although this may be due to other factors there is evidence that the risk of CAD increases with decreasing vitamin D levels[202] and biological evidence that vitamin D in the form of calcitriol has vaso-protective properties.

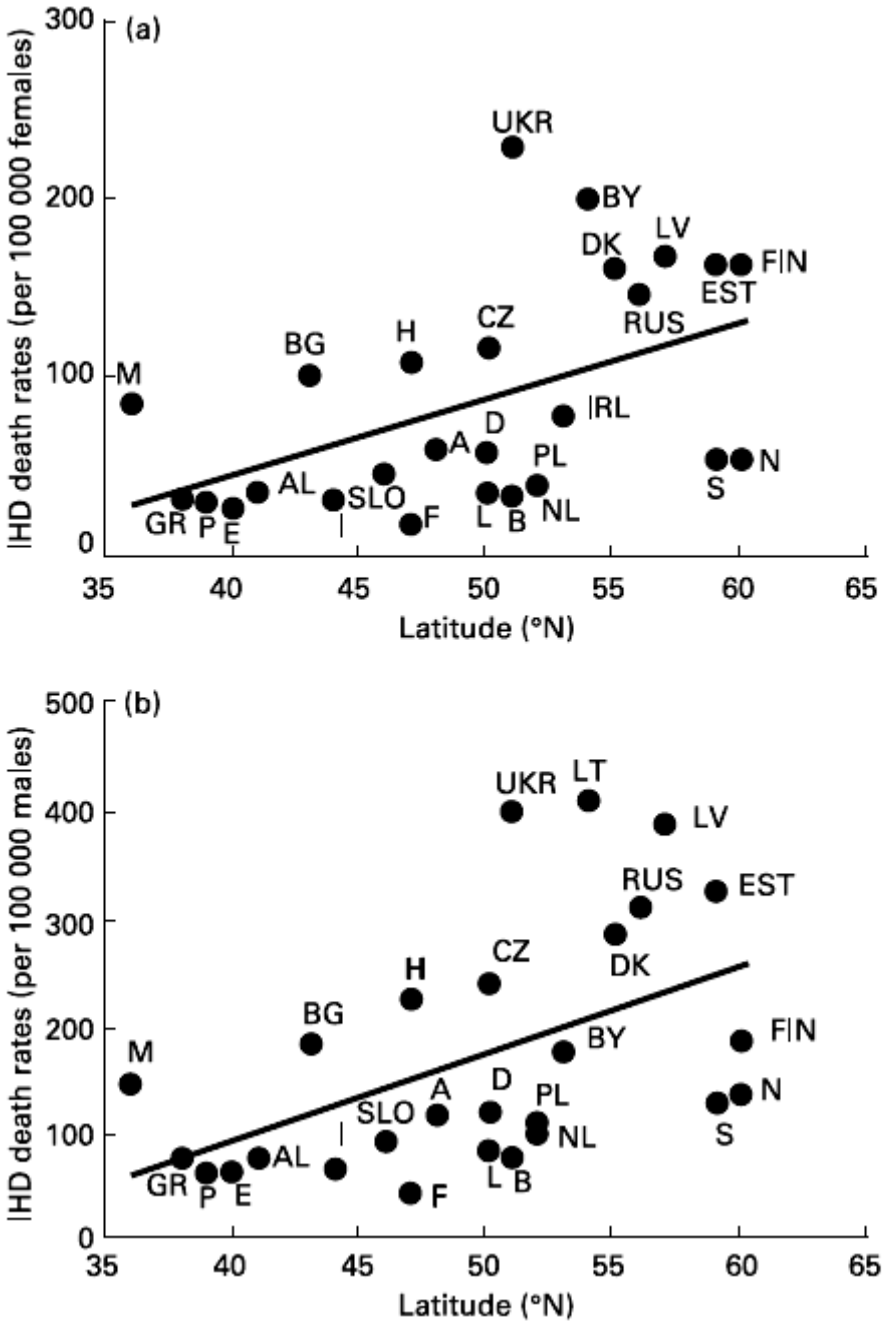

Figure 1-5: Association between geographic latitude and ischaemic heart disease (IHD) death rates in (a) females (r =0.49; P< 0·01) and (b) males (r=0.51; P<0·01) of different European countries. A, Austria; AL, Albania; B, Belgium; BG, Bulgaria; BY, Belarus; CZ, Czech; D, Germany; DK, Denmark; E, Spain; EST, Estonia; F, France; FIN, Finland; GR, Greece; H, Hungary; I, Italy; L, Luxembourg; LT, Lithuania; LV, Latvia; M, Malta; N, Norway; NL, Netherlands; P, Portugal; PL, Poland; S, Sweden; SLO, Slovenia; RUS, Russia; UKR, Ukraine[198].

1.5.2 Vitamin D metabolism

Vitamin D is a broad term which describes a group of fat soluble vitamins that are responsible for the intestinal absorption of calcium and phosphate. Vitamin D can be absorbed as cholecalciferol or ergocalciferol from the gut or can be synthesised from cholesterol in the skin with adequate sun exposure. Many different forms of vitamin D are found in different organs of the body: in the liver vitamin D is converted to calcidiol (25 hydroxyvitamin D) then part of the calcidiol is converted to calcitriol in the kidneys (Figure 1-6). Calcitriol is the biologically active form of vitamin D[203] which acts as a hormone to regulate serum concentrations of calcium and phosphate. Outside the kidneys calcitriol affects the cell cycle, inflammation and neuromuscular function[204-206].

1.5.3 Biological mechanism of vitamin D

Vitamin D plays a crucial role in maintaining healthy bones by maintaining calcium and phosphate homeostasis. Calcium and phosphate homeostasis is essential for the deposition of bone mineral and an imbalance will affect skeletal integrity. Vitamin D stimulates the absorption of calcium and phosphate in the intestines and is also responsible for the reabsorption of bone calcium and phosphate as well as reabsorption of renal calcium and phosphate. Vitamin D deficiency can lead to nutritional rickets[207]

1.5.4 Vitamin D as a vaso-protective agent

25-hydroxyvitamin D (25OHD) otherwise known as calcitriol has been shown to protect the body against atherosclerotic lesions and vascular calcification. A key step in atherosclerotic lesion development is the proliferation of SMCs and calcitriol prevents the proliferation of vascular SMCs by allowing large amounts of calcium into the cytoplasm of these cells[208]. Vascular calcification occurs in mature atherosclerotic lesions and is inhibited by calcitriol. Calcitriol increases the production of matrix Gla protein (*MGP*)[209] in vascular SMCs[210, 211] where *MGP* is known to inhibit vascular calcification[210, 212].

Chronic inflammation is also associated with the development of atherosclerosis and vitamin D has anti-inflammatory properties. *IL-6* and *TNF- α* are chief physiological stimulants of C-reactive protein (CRP), which can serve as an indicator of inflammatory processes. Increasing levels of vitamin D suppresses the release of *TNF-α* and *IL*-6 in a

dose dependent manner[213]. Vitamin D also increases the synthesis of *IL10*, an anti-inflammatory cytokine[214]. IL-10 deficiency has been associated with severe atherosclerosis in experimental animals[215].

Calcitriol levels inversely correlate with serum parathyroid hormone (*PTH*) (Figure 1-6), which has been linked to CAD risk. Individuals with vitamin D deficiency have elevated levels of *PTH* that are toxic to the vascular system. *PTH* has been linked to elevated cardiac contractility, vascular calcification and has been linked to chronic atherosclerosis via insulin resistance[216].

There is also evidence that vitamin D reduces blood pressure which in turn reduces the risk of CAD. Vitamin D acts on blood pressure through the renin-angiotensin system[217] therefore playing a central role in the regulation of blood pressure and electrolyte homeostasis[218]. Supplementation with Calcitriol reduces blood pressure, plasma renin activity, and angiotensin II levels[219, 220]. UVB radiation, key in the synthesis of 25(OH) D-serum in the skin, was shown to lower blood pressure in patients with mild essential hypertension[221, 222].

With so much evidence to support the role of increased vitamin D levels with improved vascular health it follows that decreased vitamin D levels are likely to increase risk of CAD through their action on a variety of cells in the vasculature and organs. The relationship between vitamin D and CAD could be confounded by the lack of outdoor activity observed in patients with CAD. Randomised control trials (RCT) are used to assess causal relationships between an environmental exposure and a disease outcome. In RCTs it is possible to account for the majority of confounding factors and to identify a causal relationship if one exists[223].

The results of observational studies don't always reflect what is observed in RCTs and this is often due to confounding by unobserved lifestyle factors or may be due to reverse causation. The participants in RCTs don't always represent the general population and are not powered to detect exposure that takes years to develop. Large observational studies are able to include enough people to represent the general population and observational studies run for long periods of time[223] thus allowing for adequate exposure. Another approach that can be used to mimic RCTs in observational data is Mendelian randomisation. This approach uses variants known to be predictive

of the exposure to determine the causal relationship between the exposure and the outcome.

1.5.5 The genetics of Vitamin D

Recent genome wide association studies have identified 4 variants that were associated with decreased vitamin D levels. The single nucleotide polymorphisms are located within or near 4 genes related to vitamin D metabolism: rs2282679 in vitamin D binding protein (*GC)*; rs12785878 in 7-dehydrocholesterol reductase (*DHCR7*); rs10741657 in cytochrome P450 2R1 (*CYP2R1*) and rs6013897 in *CYP24A1* were associated with differences in 25OHD levels[2, 6]. *GC* is responsible for binding vitamin D and transporting it through the blood stream; *DHCR7* metabolises 7-Dehydrocholesterol to cholesterol reducing the vitamin D precursor levels in the skin, *CYP2R1* metabolises cholecalciferol to calcidiol in the liver and *CYP24A1* is responsible for inactivating vitamin D from the kidneys (Figure 1-6). These SNPs may be used in a Mendelian randomisation study design to assess the relationship between decreasing vitamin D levels and CAD.

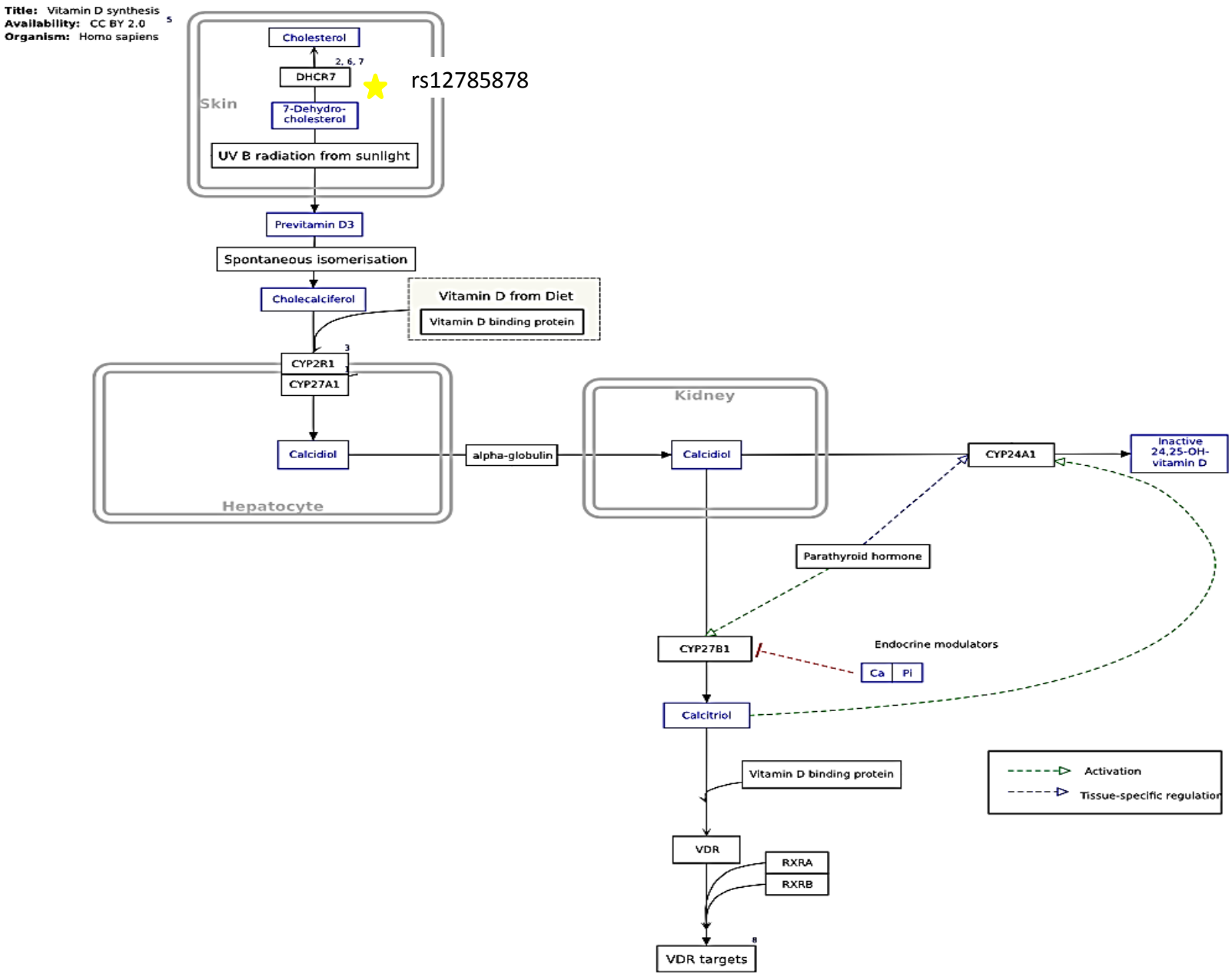

Figure 1-6: Vitamin D synthesis pathway[1-7]. This figure was generated using WikiPathways.

1.6 Mendelian Randomisation

Mendelian randomisation is an instrumental variable approach that is well suited to inferring causality when unobserved confounding is believed to be likely. It uses a well understood genetic variant known to be associated with the exposure but without a direct effect on the disease as the instrument[223]. Genotypes are assigned randomly at birth due to meiosis so the genotypes are assigned independent of disease status or the unobserved confounder. Essentially this is a randomised control trial that runs for the life time of the individual[223].

The genetic variants are used as the instrumental variable in this case would be SNPs associated with vitamin D deficiency. The association between the SNPs and CAD would provide evidence for or against a causal relationship between low vitamin D exposure and the increased risk of CAD (Figure 1-7).

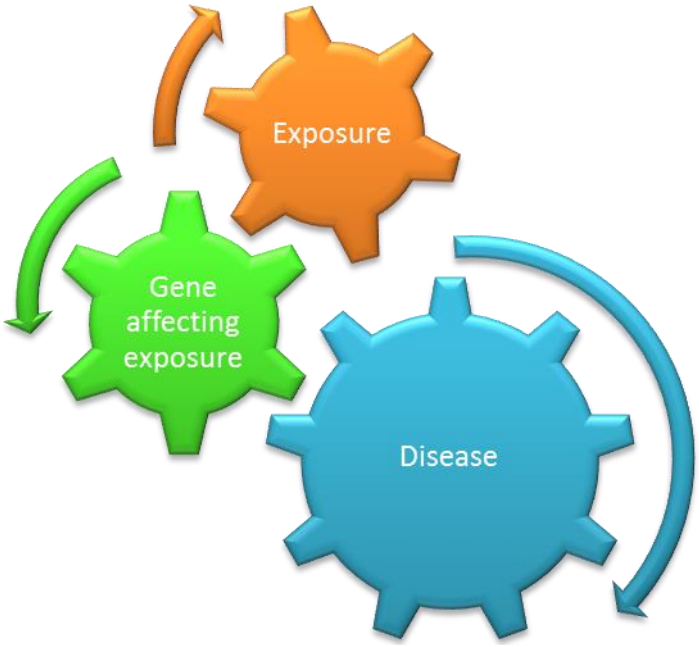

Figure 1-7: Mendelian randomisation is an instrumental variable approach which uses genes to determine the relationship between an exposure and a disease.

1.7 Rationale for the thesis

This thesis encompasses 5 specific studies with the broad aim of identifying genetic variants associated with CVD. At present, only 10% of the total heritability of CAD is explained by the known variants, which indicates that there are additional genetic variants that have not yet

been identified. To date only two genome significant variants have been identified for LEAD and no stratified analyses have been carried out leaving this phenotype largely under investigated in large scale genetic studies. None of the current CAD or LEAD associated variants overlaps with T2D associated loci, which is interesting as diabetes is a large predictor of CAD and LEAD. The Genetics of Diabetes and Audit research in Tayside Scotland (GoDARTS) is a partner in a large EU funded innovative medicines initiative (IMI) known as SUMMIT. The SUrrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools (SUMMIT) study (http://www.imi-summit.eu/) includes 19 academic and 6 industry partners across Europe with the main purpose of identifying genetic risk factors for chronic diabetic complications such as CAD and LEAD.

While investigating determinants of CVD in T2D populations may explain some of the missing heritability for CAD and LEAD, it is still useful to identify loci that may be globally associated with CVD. Variants identified from large genetic studies can also be used to determine causal relationships in observational studies. The causal relationship between vitamin D and CAD has not been analysed in a way that is free from confounding and the true relationship remains unexplained.

We hypothesised that there are genetic determinants that are associated with cardiovascular disease in patients with diabetes and in non-diabetic individuals. Therefore the main aims of this thesis are:

- To develop methods to identify individuals with CAD, IS and LEAD from the electronic medical record linked to GoDARTS for genetic association studies
- To identify new variants associated with CAD through large scale genetic studies
- To identify variants associated with CAD in T2D populations
- To determine the causal relationship between vitamin D and CAD
- To identify new variants associated with LEAD

Chapter 2: Methods

2.1 Description of databases

The Health Informatics Centre (HIC) in partnership with the University of Dundee, National Health Service Tayside and the information services division of national services [128] provides researchers and others with information derived from person-specific datasets. These datasets are mainly derived from data held by University of Dundee and the National Health Service and are anonymized in accordance with the Standard Operating Procedures approved by the Caldicott Guardians. In Scotland every person registered with a medical practitioner is assigned a Community Health Number[224]. This is a unique 10 digit identification number that is linked to information on address, postcode, medical practitioner registration status, deceased person and date of death. This information is held by the Tayside Health Board for the entire Tayside population. Tayside also uses this number as the patient identifier in all health care activities from primary to tertiary care thus allowing for the record-linkage of datasets.

2.1.1 DARTS

Diabetes Audit and Research in Tayside Scotland (DARTS) database includes information of all patients with diabetes in Tayside. Individuals with diabetes were identified from hospital records[225]. The database was validated against general practice records and was confirmed to be robust. The methodology used was shown to be more sensitive than general practice alone at identifying individuals with diabetes[225]. There are record linked data available for individuals within the DARTS cohort.

2.1.2 Go-DARTS

GoDARTS comprises 17,602 participants enrolled between December 1998 and May 2009 in which there are approximately equal numbers of participants without T2D (N=7773) and with T2D (N=9829). Participants with T2D were identified for enrolment through DARTS - a comprehensive and well validated region-wide clinical information system for diabetes that incorporates multiple clinical data sources[225, 226]. Age and sex matched diabetes-free participants were identified in populations within the region of Tayside from general practice records.

Relevant clinical data for all GoDARTS participants are drawn from electronic records of hospital admissions (Scottish Morbidity Register, SMR01), deaths (General Registry Office, GRO), biochemical tests and dispensed drug prescriptions, available for the Tayside region. Data are available from 1980 until present for the SMR hospital admissions data; from 1998 for deaths from the GRO; from 1993 until present for prescriptions; and from 1980 until present for biochemical tests.

The GoDARTS study was approved by the Tayside Committee for Medical Research Ethics and written, informed consent was obtained from each participant. A single sample of blood was collected for DNA extraction and genotyping, and the participant was assigned a unique anonymized system identifier. Baseline characteristics were recorded at time of recruitment for all participants[226].

2.1.3 CHI master index

This is a demography database that contains one entry per study individual. This file includes information on the date of birth, ethnicity, sex and date of recruitment to the Go-DARTS study.

2.1.4 SMR01

Scottish morbidity record 01 is a record of acute hospital admissions in Tayside and Fife, Scotland. The database consists of one line per patient admission that includes a date of admission, one principal diagnostic field and five additional diagnostic fields. This database also includes admissions for hospital procedures that include one principal procedure field and eight additional procedural fields. The hospital admissions are classified according to the International Classification of Diseases (ICD) 9th and 10th versions. Procedures are classified according to the Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures 3rd and 4th revisions.

2.1.5 GRO death certification

General Register Office (GRO) – Death certification database is a record of the date of death and cause of death. Deaths in Tayside have been electronically recorded since 1989 and the database includes a principal cause of death field and ten additional related causes of death fields. Cause of death is classified according to ICD9 and ICD10 codes.

2.1.6 Laboratory data

The Tayside laboratory systems record all tests performed in surgeries, clinics and hospitals that have been sent to the Tayside laboratories for processing. Clinical laboratory data are available from 1992. The database contains biochemical, haematology, microbiology, virology and serology laboratory results and reports.

2.1.7 Prescribing data

HIC provides complete data for prescriptions dispensed in Tayside from 1993. Prescriptions dispensed between 1993-2004 were recorded as scanned paper prescriptions analysed with purpose written software. Since late 2004 all prescriptions were obtained in electronic format from the Practitioner Services Division (PSD). The PSD are responsible for the processing and pricing of all prescriptions in Scotland. The prescriptions recorded include all those dispensed in community pharmacies, dispensing doctors, and a small number of specialist appliance suppliers. Hospital prescriptions are included if they were dispensed in the community.

Individual drug prescriptions are linked to an individual CHI number and are explicit on the name of the drug, date of prescription, amount dispensed as well as dosing instructions. Drugs are identified by name and individual drug codes linked to the British National Formulary (BNF).

2.1.8 Vascular Laboratories Data

2.1.8.1 Segmental Pressures Data

The Segmental pressures data includes a subset of patients that have been referred to the vascular laboratories for a LEAD assessment. Measurements are taken to record the ankle brachial pressure indexes (ABI) on both the left and right side of the body, toe brachial pressure indexes (TBI) on both left and right size, exercise tests and records of claudication.

2.1.9 Electronic stroke information system for Tayside

The electronic stroke information system for Tayside (ESIST) has been designed as a research database for the study of stroke in Tayside. ESIST contains information on individuals who have been admitted to the stroke ward with a stroke from 2006. The database contains detailed information on the diagnosis that includes whether the stroke was a trans-ischaemic attack, an ischaemic stroke or a haemorrhagic stroke. The strokes

were then identified by subtype: Cardioembolism; large artery atherosclerosis; small vessel occlusion; stroke of other determined aetiology and stroke of undetermined aetiology. Information on blood pressure, smoking status, duration of hospital stay and discharge medication are available for each stroke event recorded in ESIST.

2.2 Electronic medical records data manipulation

The electronic medical records (EMR) are supplied as flat text files from which the relevant data are extracted and combined in forms suitable for statistical analyses. Phenotypes were derived from the EMR and all covariates were extracted from the EMR using the R statistics package[227].

2.3 Statistical analysis

Statistical analyses were conducted in R[227] using a variety of statistical packages available for the R statistic programs. The theory of the main regression analyses are given below.

2.3.1 Multiple Regression

Multiple linear regression is used to determine the relationship between a dependent variable and a number of predictor variables. One of the assumptions made is that there is a linear relationship between the dependent and independent variables so the following equation is used:

$$Y=a+b_1X_1+b_2X_2+\ldots+b_nX_n+e$$

Where Y is the dependent variable, X is the predictor variable and b is the effect that the individual predictors have on determining the value of Y. Correlation between the dependent and independent variables can be demonstrated in a graph where a is the intercept and b is the slope of the line. The error term (e) represents a combined effect of the omitted variables. This is the variance in the model that cannot be explained by the independent variables included in the model[228].

While performing multiple regression the following assumptions are made: that the residuals in the model are normally distributed; error terms are constant and do not depend on the value of the independent variable and that independent variables are not correlated. Residuals are computed by subtracting the observed values of the dependent variable from those that are predicted by the model[228]

Linear regression is applied in cases where the independent variable is continuous and the b values can be interpreted as the expected change in y for one-unit change in the dependent variable if all other dependent variables are held constant. Logistic regression is applied in cases where the dependent variable is binary i.e. has two distinct values such as the occurrence of an event. The b values or regression coefficients describe the size of the contribution each step in the independent variable makes in predicting the identity of the dependent variable i.e. 1 or 0.

2.3.2 Survival analysis using a Cox's proportional hazards model

Survival analysis or a Cox's regression is often used in epidemiological studies to model the independent variables that determine the dependent variable which in this case is time to failure. In epidemiological studies an observation time is defined, usually time from patient recruitment until an outcome, death or end of the observation period. Each individual has a value for the observation time and a binary measure for outcome 1/0. To make up the Survivor function the survival time is broken up into intervals and for each time interval the proportion of individuals that have not failed and go on to enter a time period is measured. The number of cases that have had an outcome can be measured and the number of individuals that were censored for that interval can also be computed[229].

The Cox's proportional hazard model can be applied to any distribution of the survivor function, the only assumptions that must be satisfied is that there are proportional hazards i.e. the estimation of the hazard function from independent variables does not depend on time and that there is a log-linear relationship between the independent variables and the underlying hazard function.

$$h((t), (z_1, z_2, ..., z_m)) = h_0(t)*\exp(b_1*z_1 + ... + b_m*z_m)$$

Where the h(t) is the resultant hazard function given the values of the dependent variables for the cases in a time interval and $h_0(t)$ is the baseline hazard function when all the dependent variables are equal to zero[229].

Survival analysis is applied when survival time is of interest especially in terms of medical treatment. The method is tolerant of varying observation periods which is useful when patient recruitment has occurred at multiple time points so the observation periods are not

equal for all individuals in the study. Individuals who did not survive until the end of the study and have missing data for that time point wouldn't be included in the study group for a multiple regression allowing for survival bias in multiple regression methods[229].

2.4 High Density Array Genotyping Data

Samples, which were genotyped on the Genome-Wide Human SNP Array 6.0, were processed at Affymetrixs' service laboratory for all samples passing Affymetrix's laboratory quality control; raw intensities were renormalized within collections using CelQuantileNorm. These normalized intensities were used to call genotypes with an updated version of the Chiamo software adapted for Affymetrix 6.0 SNP data. The sample processing for the CardioMetabochip was undertaken at Oxford University, Oxford and for the Immuno Chip processed at the Sanger Institute, Cambridge was identical to the above. The genotypes from the CardioMetabochip and the Immuno Chip were called using the Geno SNP algorithm.

2.4.1 DNA preparation and genotyping

The quality of the genomic DNA was validated using the Sequenom iPLEX assay designed to genotype four gender SNPs and 26 SNPs present on the Illumina Beadchips. DNA concentrations were quantified using a PicoGreen assay (Invitrogen) and an aliquot assayed by agarose gel electrophoresis. A DNA sample was considered to pass quality control if the DNA concentration was greater than or equal to 50 ng/µl, the DNA was not degraded, the gender assignment from the iPLEX assay matched that provided in the patient data manifest and genotypes were obtained for at least two thirds of the SNPs on the iPLEX.

2.4.2 Sample Quality Control

Genotype data quality control of the discovery samples was similar to other Wellcome trust case control consortium 2 (WTCCC2) studies published elsewhere[99, 191, 230-232].

For all individuals, we explicitly modelled the data as a mixture of 'normal' and 'outlier' individuals for each of ancestry, missing data and heterozygosity, and sex assignment. We fitted each model in a Bayesian framework and excluded individuals whose posterior probability of belonging to the outlier class was above 0.5[145]. This approach replaces the traditional concept of fixed exclusion thresholds for parameters such as call rate, heterozygosity and ancestry.

To assess relatedness among study individuals, we compared each individual with the 100 individuals they were most closely related to (on the basis of genome-wide levels of allele sharing) and used a hidden Markov model (HMM) to decide, at each position in their genome, whether the two individuals shared 0, 1 or 2 chromosomes identical by descent (IBD). This allowed a more refined assessment of the relatedness between individuals than genome-wide sharing statistics (for example, parent-child relationships can be distinguished from those of siblings). Individuals were removed from the study iteratively to ensure there was no pair of individuals with IBD>= 5%. Within each pair of putatively related individuals, the individual with more missing genotypes was removed.

2.4.3 Affymetrix 6.0 SNP genotyping array

4000 diabetic cases were genotyped on the Affymetrix 6.0 SNP genotyping array that includes 1 000 000 SNPs. These individuals were specifically chosen for genotyping as they had all gone on to receive statins after recruitment to GoDARTS.

2.4.4 Illumina Omni-express array

4000 diabetic cases were genotyped on the Illumina Omni-express array which consists of ~700K SNPs selected from Hap Map 1-3 for SNPs with a MAF greater than 5%. The array was designed by selecting tag SNPs to serve as a proxy for a number of others SNPs across the genome. This approach allows for the broadest selection of maximally informative markers, resulting in genome-wide coverage of both common and rare variants.

2.4.5 CardioMetabochip

7,500 diabetic cases and controls were genotyped on the CardioMetabochip chip which consists of 196,725 SNPs. The loci on this custom array were chosen by Body Fat Percentage, CARDIoGRAM (coronary artery disease and myocardial infarction), DIAGRAM [233], GIANT (anthropometric traits), Global Lipids Genetics (lipids), HaemGen (haematological measures), ICBP (blood pressure), MAGIC (glucose and insulin), and QT- IGC (QT interval) GWAS meta-analysis consortia. Loci were selected based on previous findings from GWAS, individual selections from consortia and earlier genetic studies[234]

2.4.6 Immuno Chip

The Immuno Chip is similar in design to the CardioMetabochip Chip. It was designed by consortia specifically to target loci associated with 12 immunologically related human

diseases. It is an Illumina Infinium HD genotyping array designed to integrate relevant 1000Genomes data with disease-specific resequencing data and known immune-mediated disease loci that were identified by common variant GWAS.

2.5 Manipulation of the genetic data

High density SNP array data can be stored in a variety of formats. For the purposes of this thesis two file formats were used: the Chiamo[47] and the PLINK binary PED[235] formats. PLINK binary files were used when cleaning and analysing directly genotyped data. PLINK has a number of functions which were used to run the quality control checks for both sample and genotyping data described above. PLINK was used to prepare the input files for haplotype inference and subsequent genotype imputation.

2.6 Genotype imputation

Imputation takes place in two stages: the first is the estimation of haplotypes from the study population and the second is the imputation of genotypes by comparing study haplotypes to reference panel haplotypes.

2.6.1 Haplotype inference

There exist many methods for haplotype inference in unrelated populations[236] that estimate haplotypes with varying accuracy. The segmented haplotype estimation and imputation tool (SHAPE-IT, http://www.shapeit.fr/)[237] is the recommended method to estimate haplotypes for downstream imputation with IMPUTEv2[238]. The method is highly accurate and computational light when compared with other available methods[239] and is particularly suited to populations that contain high linkage disequilibrium in their genomes such as Caucasians[237]. The inference of haplotypes is computed in a similar way to Phase v2[240] where all possible haplotypes are estimated from the available genotype set with an associated probability (Figure 2-1a). Since the number of haplotypes increases exponentially with the addition of more genotypes, there need to be methods to reduce the haplotype set to the most likely haplotype set.

SHAPE-IT reduces the complexity of previous methods by collapsing the haplotype set into a graph structure and then performing hidden Markov model (HMM) calculations on the graph (Figure 2-1). So there is a set of all the possible haplotypes given the genotype set. The total haplotype set is then split into smaller segments that are disjoint (Hg) (Figure 2-

1a). The nodes of each disjoint set are represented as a single marker that can have one of two alleles; the possible haplotypes in the disjoint sets are then estimated. This new set is compared to all the possible haplotypes given the genotype set then each edge is weighted by each haplotype in the total estimated set that traverse that edge (Figure 2-1a).

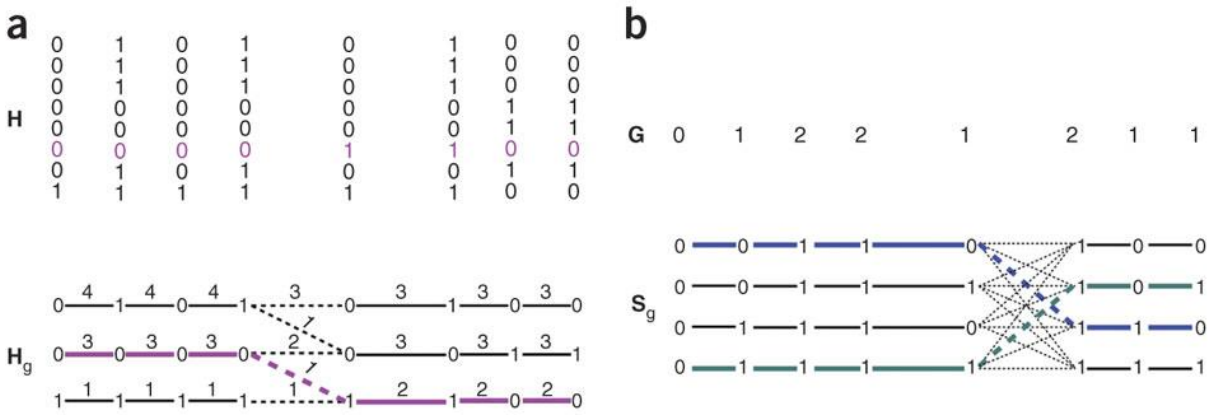

Figure 2-1: Illustration of the SHAPE-IT model and the associated graphs in a simplified example. (a, b) In this example, H contains K = 8 haplotypes (rows in a) and the individual's genotype G contains four heterozygous SNPs (b), both defined over M = 8 markers (columns). In a, we illustrate how the graph Hg is built by splitting the haplotypes of H between markers 4 and 5, resulting in two segments that each contain J = 3 distinct haplotypes. The nodes of the graph are labelled either with allele 1 or allele 0. Each edge is weighted by the number of haplotypes in H that traverse it. A haplotype of H and its corresponding path in Hg is illustrated in magenta. In b, we illustrate how the graph Sg is built by making two segments of five and three SNP markers, each one containing two heterozygous markers in G (represented as state 1; state 0 and 2 are wild type and homozygous, respectively). Each segment has four possible haplotypes compatible with G. A pair of paths in Sg compatible with G is coloured blue and green[239].

This process is iterative where the haplotype set becomes progressively smaller based on the transition probabilities between segments and the weights at the edges of the graph. Segments are chosen based on the number of heterozygous genotypes in that segment.

2.6.2 Imputation with IMPUTEv2

IMPUTEv2 (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) uses the estimated haplotypes from SHAPE-IT to impute genotypes from the haplotype set[238]. Impute2

compares the study population haplotypes with up to two reference panels for the imputation of missing genotypes in the study population. Alleles are imputed into the study population by running a forward-backward algorithm to impute missing alleles with a certain probability[238]. The two sets of haplotypes are compared to each other and missing alleles are imputed into the study panel from the reference panel with certain probability. Certain SNPs will be found in haplotypes together so if one SNP is not present in the study panel but the haplotypes match we can impute an allele for that SNP with a certain probability given the alleles that are present in the reference panel and the alleles that are given in the study panel. So if there is good coverage of a haplotype on a particular chip and the haplotype is present in the reference panel the alleles will be imputed fairly accurately. If haplotypes are sparsely covered or SNPs are not in linkage disequilibrium with any other SNPs to form haplotypes then the alleles may be imputed with low confidence in their accuracy or may be missing all together. Given that we assume both sets of haplotypes are sampled from populations in Hardy-Weinberg equilibrium the allelic probabilities can be converted to genotypic probabilities[238].

2.7 Cluster computing

The manipulation of these large genotype files and the imputation procedures are computationally intensive. Parallel computing on a high performance cluster (HPC) of computers was used to analyse and impute the genetic data. The HPC is managed by a Sun Grid Engine which is responsible for accepting jobs from users, scheduling and distributing jobs to cores within the HPC. SHAPE-IT supports threading of multiple cores so that they can be used as one processor to perform the pre-phasing steps chromosome by chromosome. Specific Perl programs were written to run the prephasing steps on the HPC using the multi-threading capabilities of SHAPE-IT. Each chromosome was imputed by dividing the imputation intervals across 5Mb chunks of each chromosome. Each chunk was distributed to a separate core and specific Perl programs were written to run each of the imputation steps. The resulting files were combined into single chromosomes using bash scripting. Imputed files were also analysed chromosome by chromosome on the HPC. The analysis programs used for directly and imputed genotypes are described below.

2.8 Genetic associations

Genotypes of directly typed SNPs are assigned using a threshold method at a probability of 0.9. SNPs can be modelled in logistic and linear regressions on a log-additive scale. Similar to non-genetic covariates the effect of the SNP is modelled by steps from one genotype to another. SNPs are coded 0 for homozygotes of the non-effect allele, 1 for one copy of the effect allele in a heterozygote and 2 for a homozygote of the effect allele so the effect of each step translates to the effect of an additional copy of the effect allele on the outcome. Imputed SNPs are more complex to model as the three genotypes are estimated with some probability and the imputed data are assumed to be 'missing' in that they are not observed. Models of association for imputed genotypes need to take into account genotype uncertainty.

SNPTESTv2 (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/old/snptest.html) takes genotype uncertainty into account by using a missing data likelihood score test. Since SNPs may be imputed with varying certainty a threshold method may lead to a high proportion of missing data. It is possible to sum the probabilities of each genotype across individuals for a SNP and use all the available data in the association analysis. For a particular SNP data may be missing in some individuals and not in others so the data can be partitioned into observed and missing data. A score test can be used to estimate the likelihood of the observed data given the full set of data and taking into account the missing data[241].

2.9 Calculation genotypic scores

The genotypes of the SNPs were coded as 0 - no effect alleles present; 1 - one effect allele present and 2 - two effect alleles present. Effect alleles were defined as those reported to raise levels of quantitative traits or risk of diabetes. Where the reported SNP was not typed or imputed a suitable proxy was identified. A genotype score for each trait was calculated for every participant based on the number of effect alleles the participant had for each SNP associated with a trait. Each SNP was weighted by the published per affect allele increase in quantitative trait or binary outcome (allele effect estimates): GS = SNP1 (0/1/2)*β_{Effect_allele} + SNP2 (0/1/2)*β_{Effect_allele} … + SNP_{ij} (0/1/2)*$\beta_{Effect\_allele.}$ If a genotype was missing for an individual, a score value was imputed based on the probability of a homozygous effect allele

genotype for that individual multiplied by the effect allele weight: $2*MAF_{effect\_allele}*\beta_{Effect\_allele}$.

2.9.1 Assessing the discriminatory power of genetic risk score to classify disease status

If genetic risk scores are to be applied to predicting disease status we need methods to assess the discriminatory power of the genetic risk score to predict disease outcome. Since the outcome is binary we need methods that can specifically be applied to binary outcomes and the most commonly derived metrics are the R squared, concordance (C) statistic and more recently net reclassification improvement.

The coefficient of determination or the R-squared value is a measure of the "goodness of fit" for a regression model and was developed for linear regression models. New methods have been developed to calculate the R-squared for logistic regression models. The value of R-squared ranges between 0 and 1 and the higher the value of R-squared the better the model is at predicting the outcome[242]. The C statistic is another measure of how well a model predicts the disease outcome.

The C statistic for binary outcomes is the area under the receiver operating curve and takes on values between 0.5 and 1. The receiver operating curve is a graphical plot that illustrates the performance of a binary classifier; in this case it could be used to compare the predictive power of different logistic regression models to each other. The closer the c-statistic is to 1 the larger the area under the ROC curve which indicates a higher discriminatory power for a given model[243]. The C statistic has been criticised for being too conservative for a meaningful improvement in prediction because increases can only be seen for variables that carry a very high relative risk for the disease outcome.

The net reclassification improvement (NRI) is a new method to assess the improvement of model performance offered by a new variable [244]. The method uses reclassification tables for constructed separately for diseased patients and those free of disease. Any upward movement in patients with disease is considered an improvement in classification while a downward movement is considered a worse reclassification. The opposite is true for individuals who do not suffer disease events. The improvement in classification can be quantified in patients who suffer events by adding the differences in the proportion of

individuals who move up minus those who move down. The inverse is applied to individuals who do not suffer disease events. The classification in each group can be compared by taking the difference in reclassification in the event and non-event groups. The larger the difference the more significant the improvement in disease prediction provided by the new variable[244].

2.10 Laboratory methods

2.10.1 TaqMan Genotyping

Direct typing of individual SNPs was performed using TaqMan allelic discrimination assays as supplied by Applied Biosystems (Carlsbad, CA) as "Assays on Demand", or "Assays by Design". All typing was performed in 384 well format using 10-20ng of DNA in 2ul reaction volumes using Universal TaqMan master mix (Applied Biosystems, Carlsbad, CA). Assays were plated using a DEERAC Equator GX microdispenser (Labcyte, Sunnyvale, CA), and thermal cycling was performed in a H2OBIT high throughput thermal cycler (KBiosystems, Basildon, Essex). End point fluorescence was measured and genotypes were called using an ABI 7900HT sequence detection system (Applied Biosystems, Carlsbad, CA).

2.11 Meta-analysis

In a meta-analysis summary statistics such as effect estimates and the variance of the effect estimates from individual studies are combined. This process allows us to estimate the effects of SNPs across any number of studies without the need to access individual level data. Summary statistics can be combined using either a fixed effects or random effects model[191, 245-247].

2.11.1 Fixed effects meta-analysis

The fixed effects model assumes that the effect sizes are the same or fixed across studies. Effects are commonly combined using inverse-variance weighted effect size estimate and the weighted sum of z scores.

$$\beta meta = \frac{\sum Wi\beta i}{\sum Wi}$$

Where $\mathrm{Wi} = 1/(\mathrm{SE})^2$ and the $SE_{meta} = \sqrt{\sum Wi}^{-1}$

$$Z_{FE} = \frac{Bmeta}{SEmeta}$$

A p value can be calculated from the z score assuming a two-sided test

$$P_{FE} = 2\Phi(-|Z_{FE}|)$$

Where Φ is the cumulative density function of the standard normal distribution.

191, 245, 246

2.11.2 Random effects meta-analysis

The random effects model assumes that the effect sizes are sampled from a probability distribution that has a variance τ^2 that can be estimated using various approaches[246, 248] such as the method of maximum likelihood and restricted maximum likelihood[248]. The estimated between study variance τ^2 hat is included in the calculation of the effect size estimate where

$$\beta = \frac{\sum(Wi + \hat{\tau}^2)^{-1}\beta\, i}{\sum(Wi + \hat{\tau}^2)^{-1}}$$

$$SE(\beta) = \sqrt{\sum (Wi + \hat{\tau}^2)^{-1}}^{-1}$$

And the z score can be calculated as follows:

$$ZRE = \frac{\beta}{SE(\beta)}$$

The p value can then be calculated as $P_{RE} = 2\Phi(-|Z_{RE}|)$

The current RE model assumes that there is heterogeneity under the null hypothesis however there should be no heterogeneity under the null hypothesis as all effect sizes are equal to zero. This assumption leads to overly conservative p values[246].

2.11.3 RE approach Han and Eskin

The approach proposed by Han and Eskin partitions the estimation off the effects and the significance of the association into two parts. First the effect size and variance of the effect

size are estimated using the traditional random effects approach. To estimate the significance of the result a likelihood ratio test is applied to the effect estimate (ϒ)[246]. Where the likelihood of the effect estimate is maximised through a number of iterations and $\hat{\tau}^2$ is also estimated through a series of iterations. The calculation of the significance of the test statistic is complex as $\hat{\tau}^2$ is non-negative (there cannot be negative variance in an effect estimate) and follows a half normal distribution while the effect estimate (ϒ) is unrestricted and can be either positive or negative. This means that the test statistic follows a 1df and a 2df chi square distribution simultaneously. The authors provide tables for corresponding p values under different conditions allowing for unequal sample sizes and small sample sizes. Therefore no heterogeneity is assumed under the null hypothesis but is taken into account when estimating the combined effect size and variance. This random effects method has been implemented in MetaSoft (http://genetics.cs.ucla.edu/meta/)[246].

2.11.4 Heterogeneity statistics I and Q

Cochran's Q statistic is computed by summing the variation of each study's effect estimate from the overall effect estimate and weighting each study contribution by its inverse variance. The null hypothesis is that there is no heterogeneity in the effect sizes. The Q statistic follows a chi-square distribution with K-1 degrees of freedom where K is the number of studies included in the meta-analysis. The Q statistic has poorer power to detect true heterogeneity when the sample sizes are small.

The I^2 is another statistic used to estimate the between study heterogeneity and is related to the Q statistic. The I^2 described the total variation across studies that is due to heterogeneity rather than chance.

$$I^2=\frac{100\%*(Q-df)}{Q}$$

Where Q is Cochran's Q. I^2 differs from Q as it not only indicates that there is heterogeneity present, it also indicates how much heterogeneity there is: 25% (I^2 = 25), 50% (I^2 =50), and 75% (I^2=75) would mean low, medium, and high heterogeneity, respectively.[249, 250].

2.12 Testing for heterogeneity between estimates

Interaction can be measured simply as the difference between two effect estimates. The two estimates have to be independently estimated and not obtained from the same individuals[251]. Effect estimates and their standard errors are required from both groups to calculate a z score that can be used to test for an interaction. The difference between the two effects can be calculated as follows[251]:

$$d = E1 - E2$$

Where E1 and E2 are effects estimated from two separate studies. The standard error of the difference is the square root of the sum of the variance of each effect estimate:

$$SE(d)=\sqrt{SE(E1)^2 + SE(E2)^2}$$

Similar to the fixed effects method a z score can be calculated as:

$$Z=\frac{d}{SE(d)}$$

Z can be compared to a standard normal distribution to get a p value. A significant p value indicates that d is sufficiently different to zero i.e. no difference in the effect estimates[251].

2.13 Tests for stratum specific effects

GWAMA (http://www.well.ox.ac.uk/gwama/)[252] has the option to analyse sex specific effects in a meta-analysis using a fixed-effects model.

$$\beta Mj = \frac{\sum KiWij\beta ij}{\sum Wij} \quad \beta Fj = \frac{\sum(1-Ki)Wi\beta i}{\sum(1-Ki)Wi}$$

Where K denotes the sex of the ith GWAS and is equal to 1 if the study is male-specific and 0 if the study is female specific. Where $W_{Mj} = [Ki*1/(SE)^2]^{-1}$ is the variance of male specific effect estimates and $W_{Fj} = [(1-Ki)*1/(SE)^2]^{-1}$ is the variance of the of the effect estimate in females. We can then calculate chi-square statistics for each gender by:

$x^2_{Mj} = [\sum_i K_i/S_{ij}^2]$ and $x^2_{Fj} = [\sum_i K_i/S_{ij}^2]^{-1}$ where each chi-square distribution with one degree of freedom. Cochran's Q statistic can also be computed for each gender stratum by:

$Q_{Mj} = \sum_i \frac{K_i(b_{ij} - B_{Mj})}{s_{ij}^2}$ for men and $Q_{Fj} = \sum_i \frac{(1-K_i)(b_{ij} - B_{Fj})}{s_{ij}^2}$ for women.

A sex differentiated test can then be performed by allowing for different effects in males and females given by $x^2_{Dj} = x^2_{Mj} + x^2_{Fj}$ which has a chi-square distribution with two degrees of freedom. It is then possible to test for heterogeneity of effects between sex-specific allelic effects by $x^2_{Hj} = x^2_{Dj} - x^2_{Cj}$ wherex^2_{Cj}, the chi-square statistic assuming the same allelic effects in both sexes, x^2_{Hj} has a chi square distribution with 1 degree of freedom.

The most powered analysis is achieved when there is no heterogeneity between the allelic effects in males and females for a causal variant. However, when there is heterogeneity in allelic effects the loss of power is not too large and is due to the necessary extra degree of freedom. x^2_{Dj} gives us the sex-differentiated estimate for association allowing for the heterogeneity in the allelic effects between sexes to estimate the overall association with the causal variant. x^2_{Hj} is a measure of the heterogeneity between estimates when compared to the overall meta-analysis estimate that does not take sex differentiation into account. This method can be applied to any stratified analysis for example to detect diabetic and non-diabetic specific effects on coronary artery disease.

Chapter 3: Using Electronic Medical Records to Investigate Cardiovascular Phenotype-Genotype Associations – A GoDARTS study

3.1 Introduction

There is growing interest in the use of biobanks linked to medical electronic records (EMR) for genetic association studies due to the inherent versatility of the study design[225, 253, 254]. Individuals recruited to the Genetics of Diabetes and Audit Research in Tayside Scotland (GoDARTS) are linked to routinely recorded electronic medical records (EMR) and, a large proportion of the study, to high density genotyping data. Although GoDARTS was principally set up and has been used to study the genetics of type 2 diabetes (T2D)[241], it has also been used to study pharmacogenetic interactions with metformin, commonly prescribed to treat diabetes, and statins, used to prevent and treat cardiovascular disease[255-259], it is possible to define multiple phenotypes from the extended EMR and combine them with genetic data that are common to the entire GoDARTS study. This process is more efficient than recruiting multiple disease specific populations.

Identification of individuals with cardiovascular disease from GoDARTS would provide a powerful tool to investigate the genetic determinants of cardiovascular disease. The identification process is not straightforward as EMR are kept throughout the lifetime of a patient and serve the primary purpose of providing information to health care professionals and are not for genetic association studies. Therefore algorithms to identify cardiovascular phenotypes in GoDARTS need to be optimised and well validated for applications in genetic studies.

Several genome wide association studies (GWAS) have identified loci associated with cardiovascular disease and with cardiovascular disease risk factors. The 9p21 region and the *LPA* gene have been established as important loci where 9p21 variants have been associated with CAD and IS[38, 40, 48, 161] and *LPA* variants rs107455872 and rs3798220 have been associated with CAD, IS and LEAD[50, 162]. SNPs from GWAS studies of cardiovascular related traits have been combined into genetic scores (GS) for blood pressure loci, high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C) to assess the relationship between these traits and cardiovascular outcomes[31, 260].

The main aim of this chapter was to validate GoDARTS as a discovery cohort for genetic associations in cardiovascular disease. The strategy we followed was to develop algorithms to identify individuals with and without CAD, IS and LEAD from EMR and to verify the

algorithms for IS and LEAD using alternative EMR data. To show that these phenotypes could be used in genetic association studies we compared the association of known signals in the 9p21 and *LPA* regions within our study to published data and then we also compared the associations of GS, for known cardiovascular risk factors, to published associations and amongst the EMR derived phenotypes derived in GoDARTS.

3.2 Materials and Methods

3.2.1 Coronary artery disease definition

CAD events were identified from the SMR01 and GRO death registry using International Classification of Diseases (ICD) 9 and ICD 10 codes. CAD events were defined as fatal and nonfatal records of myocardial infarction (ICD9 code 410 and ICD10 codes I21 and I22), and or, unstable angina (ICD9 code 411 and ICD10 code I20.0), and/or coronary revascularization. Coronary revascularization included percutaneous transluminal coronary angioplasty (PTCA) office of population, censuses and surveys classification of surgical operations and procedures (OPCS) 3 codes 881 and OPCS 4 codes K759, K491-K493, K498-K499, K502, K504 and K508. Coronary artery bypass surgery (CABG) was identified using OPCS 3 codes 3043, 8814 and OPCS 4 codes K401-402, K411, K413, K433, K441-K442 and K449.

3.2.2 Ischaemic stroke definition

The phenotype definition for IS was developed and validated by Flynn et al., (2010)[261]. Briefly, IS were identified from SMR01 and GRO data using the following codes: ICD 9: 434.1, 434.9, 435.9, 436.9 and ICD10: G450-4, G458, G459, I63.0-6, I63.8, I63.9, I64 and I69.3 and I69.4.

3.2.3 Lower extremity arterial disease definition

3.2.3.1 Vascular laboratory criteria

LEAD cases were identified from the vascular laboratories data based on an ankle brachial index (ABI) of less than 0.9, and or an ABI greater than 1.3, and or no record of an ABI with a corresponding brachial pressure measure either on the left or the right body side, and or an ankle systolic pressure of greater than 255mmHg. Non atherosclerotic causes of LEAD were excluded if individuals had at least two records of the following codes in the Scottish

Morbidity Register 01 (SMR01): ICD9 747.22, 237.7, 443.1, 446.0, 446.4, 446.5, 446.6, 446.7, 447.6, 710.1, 747.1, 747.64 and ICD10 I73.1, I77.6, M30.3, M31.1, M31.30, M31.4, M31.6, M34.9, Q25.1, Q25.2, Q25.3, Q27.32, Q85.00. Cases were also identified primarily based on the SMR01 data.

3.2.3.2 Hospital admissions with a LEAD diagnosis

LEAD cases were identified from the SMR01 data using codes pertaining to a diagnosis of LEAD: ICD9 440.2×, 440.3×, or 440.8×; ICD10 I73.1, I77.6, M30.3, M31.1, M31.30, M31.4, M31.6, M34.9, Q25.1, Q25.2, Q25.3, Q27.32, Q85.00. Additional cases were also identified from the SMR01 data using codes for procedures and corrective surgeries related to LEAD.

3.2.3.3 Procedures and corrective surgery related to LEAD

Procedures for lower extremity angiography with a concurrent record for a non-coronary vessel stent or a procedure code for lower extremity artery surgical and percutaneous vascular interventions were considered a positive indicator of LEAD. Lower extremity angiography: ICD9 88.48, 75710, 75711, 75712, 75716, 75717, 75718, 75630, 75631; OPCS4 L63,L71,L72, U117 AND Z38 and a concurrent non-coronary vessel stent: ICD9 39.50, 39.90, 37205, 37206, 37207, 37208, 37184, 37185, 37186 or ICD10 L76, L98. Procedure codes for lower extremity artery surgical and percutaneous vascular interventions: ICD9 : 38.18, 39.50, 39.25, 39.29, 38.08, 38.38, 38.48, 39.49; 39.56, 39.57, 39.58, 39.90, 35302, 35303, 35304, 35305, 35306, 35331, 35351, 35355, 35361, 35363, 35371, 35372, 35381, 35452, 35454, 35456, 35459, 35470, 35472, 35473, 35474, 35481, 35482, 35483, 35485, 35491, 35492, 35493, 35495, 35521, 35533, 35537, 35538, 35539, 35540, 35541, 35546, 35548, 35549, 35551, 35556, 35558, 35563, 35565, 35566, 35571, 35582, 35583, 35585, 35587, 35621, 35623, 35637, 35638, 35641, 35646, 35647, 35651, 35654, 35656, 35661, 35663, 35665, 35666, 35671, 35226, 35256, 35286, 35700, 35721, 35741, 35876, 35879, 35881, 35883, 35884, 37184, 37185, 37186, 37205, 37206, 37207, 37208; OPCS4 L51-, L52, L531, L532, L538, L539, L54, L59-, L60-, L62-, L66-, L68-, L70-, Z38-. Other reasons for the vascular surgery were excluded if the following codes were also present: ICD9 736.3×, 736.4×, 736.5, 736.6, 736.7×, 736.8×, 736.9, 735.×, 754.3×, 754.4×, 754.5×, 754.6×, 754.7×, 755.02, 755.13, 755.14, 755.3, 755.4, 755.6×, 755.8, 759.7, 759.89, 895.××, 896.××, 897.××, 820.××, 821.××, 822.××, 823.××, 824.××, 825.××, 826.××, 827.××, 828.××, 829.××, 835.××, 836. ××, 837.××, 838.××, 904.××, 928.××, 929.××, 959.6, 959.7, 996.4×, 996.66, 996.67, 996.77, 996.78; ICD10

M201, M202, M203, M204, M205, M206, M214, M216, M217, M219, M22-, M23-, M2406, M2416, M2426, M2436, M2446, M2456, M2466, M2476, M2476, M2486, M2496, M2506, M2516, M2526, M2536, M2546, M2556, M2566, M2576, M2586, M2596, M8406, M8416, M8426, M8436, M8446, M8456, M8466, M8476, M8486, M8496, Q682, Q683, Q684, Q685, Q65-, Q6-, Q692, Q702, Q703, Q704, Q72-, Q741, Q742, S72-, S730-, S75-, S77-, S78-, S797, S81-, S82-, S83-, S84-, S85-, S86-, S87-, S88-, S89-, S91-, S92-, S93-, S94-, S95-, S96-, S97-, S98-, T013, T016, T023, T025, T026, T033, T034, T043, T044, T047, T053, T054, T055, T056, T8416, T8426, T8436, T8446, T8456, T8466, T8476, T8486, T8496. Mid-thigh to mid-foot amputations were also considered positive indicators of LEAD.

3.2.2.4 Mid-thigh to mid-foot amputations

Amputations from thigh to mid foot that were not due to non-vascular causes such as accidents were used to identify LEAD cases. The following codes were used to extract records of amputation from the SMR01 data: ICD9 84.1×, 84.91, 27295, 27590, 27591, 27592, 27598, 27880, 27781, 27782, 27888, 27889, 28800, 28805 and ICD10 X09, X10, X11. Further amputations were identified from amputation data available from the Scottish Care Information – Diabetes Collaboration (SCIDC) where variables in the data indicated amputation site and side. Non vascular amputations were excluded if there was also a record for : ICD9 736.3×, 736.4×, 736.5, 736.6, 736.7×, 736.8×, 736.9, 735.×, 754.3×, 754.4×, 754.5×, 754.6×, 754.7×, 755.02, 755.13, 755.14, 755.3, 755.4, 755.6×, 755.8, 759.7, 759.89, 895.××, 896.××, 897.××, 820.××, 821.××, 822.××, 823.××, 824.××, 825.××, 826.××, 827.××, 828.××, 829.××, 835.××, 836. ××, 837.××, 838.××, 904.××, 928.××, 929.××, 959.6, 959.7, 996.4×, 996.66, 996.67, 996.77, 996.78 and ICD10 M201, M202, M203, M204, M205, M206, M214, M216, M217, M219, M22-, M23-, M2406, M2416, M2426, M2436, M2446, M2456, M2466, M2476, M2476, M2486, M2496, M2506, M2516, M2526, M2536, M2546, M2556, M2566, M2576, M2586, M2596, M8406, M8416, M8426, M8436, M8446, M8456, M8466, M8476, M8486, M8496, Q682, Q683, Q684, Q685, Q65-, Q6-, Q692, Q702, Q703, Q704, Q72-, Q741, Q742, S72-, S730-, S75-, S77-, S78-, S797, S81-, S82-, S83-, S84-, S85-, S86-, S87-, S88-, S89-, S91-, S92-, S93-, S94-, S95-, S96-, S97-, S98-, T013, T016, T023, T025, T026, T033, T034, T043, T044, T047, T053, T054, T055, T056, T8416, T8426, T8436, T8446, T8456, T8466, T8476, T8486, T8496.

3.2.2.5 Dispensed medication for Claudication

Claudication is a common symptom of LEAD therefore cases could be identified by prescription for medication used to treat claudication. These included Clisotazol, Pentoxifylline, Naftidrofuryl, Inositol nicotinate, Tymoxamine and their synonyms.

The algorithm was aligned to the LEAD case definition developed at the Mayo Clinic eMERGE (electronic MEdical Records and GEnomics) centre[262] but differed from that definition by including ICD 10, OPCS 3, OPCS4 and broader medication criteria.

3.2.4 Cardiovascular disease controls

Controls were identified from hospital admissions to be free of CAD, IS and LEAD. In addition to not meeting the case definition for CAD, IS and LEAD individuals with any lower extremity amputations regardless of site were excluded from the controls.

3.3 Sensitivity and Specificity of Ischaemic stroke and LEAD case identification algorithms

Sensitivity was calculated as the number of true positives divided by the sum of the number of true negatives and the number of true positives. Specificity was calculated as number of true negatives divided by the sum of the number of true negatives and the number of true positives.

ESIST (described under heading 2.1.9) contained records of patients admitted to the stroke ward with any type of stroke and did not include records of non-stroke morbidities. A diagnosis of stroke subtype was made by stroke clinicians and recorded in ESIST, thus a 'true' ischaemic stroke event was identified as a clinical diagnosis of ischaemic stroke and a 'negative' diagnosis of ischaemic stroke was identified as a haemorrhagic stroke or stroke of unknown aetiology recorded from ESIST. Ischaemic strokes identified from SMR and GRO records were compared to corresponding records of confirmed diagnoses of stroke subtype recorded in ESIST and sensitivity and specificity were calculated.

A clinical diagnosis of LEAD is based upon ABI readings where a normal range is considered an ABI between 0.9 and 1.2[263] and anything outside of that range is considered abnormal. In this study a 'true' diagnosis of LEAD was defined as an ABI of less than 0.9 or greater than 1.4 and a strict 'negative' diagnosis of LEAD was defined as an ABI between 1.1 and 1.2

based on ABI records in the vascular laboratories data. LEAD cases identified based on criteria that did not include the vascular laboratories data were compared to corresponding records in the vascular laboratories data. Specificity and sensitivity were calculated.

3.4 Genotype information

High density genotyping data were available for GoDARTS individuals : 4000 type 2 diabetic individuals were typed on the Affymetrix 6.0 SNP genotyping array and genotypes were imputed into this population from a reference panel of HapMap2 individuals and a panel of 6000 British individuals typed on the Illumina Human 1M dual using IMPUTE2[230, 264]; 7178 individuals were genotyped on the CardioMetabo Chip and 2500 individuals were genotyped on the Immuno Chip. A Full description of the genetic data and the imputation process are given under headings 2.4 and 2.6 respectively.

Genotypes for rs10116277, rs2383207, rs1333040, rs1333049 and rs10757274 were extracted from the genetic data available for GoDARTS and the GS were calculated from the available genetic data. Rs10744872 and rs3798220 were typed using TaqMan allelic discrimination assays in the GoDARTS individuals. TaqMan allelic discrimination is described in full under heading 2.10.1.

3.5 Statistical Methods

3.5.1 Association of known SNPs with cardiovascular phenotypes

Known variants in the 9p21 region (rs10116277, rs238207, rs1333040, rs1333049, rs10757274 and *LPA* region (rs10488572 and rs3798220) were associated with cardiovascular disease in binary logistic regressions while correcting for age, gender and T2D status using the stats package in R[227].

3.5.2 Meta-analysis of LPA SNPs in cardiovascular disease

The association of rs10455872 and rs3798220 with cardiovascular phenotypes in GoDARTS and published effect estimates[50, 162] were pooled by inverse-variance weighting[245], weighted by the number of cases contributed by each study.

3.5.3 Association between Genetic Scores for cardiovascular risk factors and their corresponding traits

SNPs identified from GWAS meta-analyses for CAD[38, 47-49, 265], lipid traits[54], blood pressure[31], T2D[266], type 1 diabetes (T1DM) [191] and fasting glucose[267] were combined into Genetic Scores (GS). The SNPs used in each of the scores are reported in the appendices 1 to 6.

3.5.4 Models used for individual traits and Genetic Scores

To assess the effect of the GS on their corresponding traits logistic and linear regressions were conducted in R v2.14.1 using the stats package[227] and were stratified by T2D status. The lipid trait GS for total cholesterol (TCHOL), LDL-C, HDL-C and triglycerides (TG) were associated with mean, statin-untreated lipid measures and were corrected for mean age and gender in a linear regression. To account for the effects of antihypertensive treatment on blood pressure measures 15mmHg was added to treated SBP measures and 10mmHg to treated DBP measures. Treated blood pressure measures were defined by a dispensed prescription for an antihypertensive medication up to 60 days prior to the blood pressure measure. The GS for DBP and SBP were associated with corrected blood pressure measures taken at study enrolment in a linear regression while correcting for age and gender. Similarly, the GS for fasting glucose (FGLU) was associated with fasting glucose while correcting for age, gender and body mass index (BMI). Fasting glucose measures were taken at study enrolment in non-diabetic individuals only as FGLU and were not available for the T2D subgroup.

GS for binary traits CAD and T2D were tested for association with their respective traits in a logistic regression. The association between the GS for CAD and CAD was corrected for age and gender and stratified by T2D status. The GS for T2D was associated with T2D while correcting for age and gender.

3.5.5 Association of genetic scores for known risk factors and corresponding traits with cardiovascular disease

BMI, T2D, history of smoking, SBP, DBP, TCHOL, LDL-C, HDL-C, TG and glucose (GLU; fasting and non-fasting) were derived from EMR records as a record on the index date or nearest record. The index date for all models was defined as the time of event for cardiovascular disease cases and time of study enrolment for cardiovascular disease free controls. Possible

treatment confounding of blood pressure measurements and lipid levels was taken into account by correcting blood pressure measures for anti-hypertensive treatment as previously described and by identifying statin treated lipid measures by a dispensed prescription for a statin up to 60 days before the lipid measurement date. T2D was defined as ever T2D.

Each GS (for SBP, DBP, T2D, T1DM, TCHOL, LDL-C, HDL-C, TG and FGLU) and corresponding risk factor were associated with a cardiovascular phenotype in a binary logistic regression using the stats package in R[227] and were stratified by T2D status. All models were corrected for age and gender. Associations with lipid GS and serum lipid measures were corrected for statin treatment by including a binary variable for treated and untreated measures in the model.

3.6 Results

3.6.1 Description of cardiovascular phenotypes and phenotype confirmation in GoDARTS

From the GoDARTS population of 17602 individuals: 2317 individuals with at least one CAD event, 701 individuals with at least one record of IS, 1881 individuals with at least one record of LEAD and 11,636 CVD controls were identified (Table 3-1). Pairwise t-tests were conducted in R v2.14.1[227] comparing cardiovascular disease cases to cardiovascular disease free controls showed that the cases were different to the controls for cardiovascular disease risk factors investigated in this study (Table 3-1).

Table 3-1: Population characteristics of each cardiovascular phenotype and the cardiovascular disease free group

Descriptive/Phenotype	**CAD mean (SD) or %**	**Ischaemic stroke mean (SD) or %**	**LEAD (SD) mean (SD) or %**	**CVD controls mean (SD) or %**
N	2317	701	1881	11636
Age	64.1 (11.5)***	72.3 (11.2)***	66.1 (11.2)***	60.0 (12.6)
Male %	68.1***	57.1*	58.9**	49.7
Type 2 diabetes %	80.9*	72.6***	82.6***	40.6
BMI (kg/m^2)	29.5 (5.1)***	29.0 (5.1)	29.6 (5.5)*	29.0 (5.8)
History of Smoking %	56.8*	63.6***	59.4***	53.9
Systolic Blood pressure (mmHg)	151.4 (19.6)***	141.9 (22.9)**	140.9 (20.9)**	137.9 (19.1)
Diastolic blood pressure (mmHg)	85.5 (11.1)***	76.6 (12.6)**	76.5 (11.4)***	78.7 (10.3)
Cholesterol (mmol/L)	5.5 (1.3)***	5.7 (1.2)***	5.2 (1.1)**	5.0 (1.1)
Low-density lipoprotein (mmol/L)	3.2 (1.0)***	3.2 (0.9)***	3.2 (1.0)***	3.2 (0.9)
High-density lipoprotein(mmol/L)	1.2 (0.8)***	1.3 (0.6)***	1.3 (0.9)***	1.4 (0.6)
Triglycerides (mmol/L)	2.3 (1.6)***	2.0 (1.5)**	2.2 (1.7)***	1.9 (1.4)
Glucose (mmol/L)	9.6 (5.2)***	9.2 (4.8)***	9.4 (5.1)***	8.2 (4.4)

* 1E-05=<p<=1E-02; **1E-10<= p<1E-05; ***p<1E-10

3.6.2 Sensitivity and Specificity analysis

The algorithm developed in this study showed high sensitivity and specificity to identify IS and LEAD cases from EMR data. The IS identification algorithm was 97.8% sensitive and 100% specific when compared to confirmed clinical diagnoses of IS. Similarly the different criteria used to identify LEAD cases were 100% sensitive and had an average specificity of 94.4% when compared to clinical diagnoses (Table 3-2).

Table 3-2: Evaluation of the lower extremity arterial disease algorithm compared with ABI defined cases and controls.

Criterion	N	In data*	Sensitivity	Specificity
Diagnosis LEAD	122	56%	100%	99.6%
Amputations	197	60%	100%	97.7%
Procedures	306	61%	100%	93.1%
Prescriptions	561	40%	100%	87.1%

* The data on segmental pressures only includes individuals that were referred to the vascular laboratories for a measurement. This indicates the percentage that was referred from each identification criterion.

3.6.3 Association of known SNPs with cardiovascular phenotypes

3.6.3.1 9p21 SNPs

The association between 9p21 variants with CAD and IS were replicated in GoDARTS and were similar to published results. The associations with IS were not statistically significant but the effect allele and effect sizes were consistent with published effects. Even though these SNPs had not been previously associated with LEAD, the associations were similar to those observed for CAD (Table 3-3).

Table 3-3: Association of SNPs on chromosome 9 and 6 associated with cardiovascular disease with cardiovascular disease phenotypes in GoDARTS.

				CAD			Ischaemic stroke			LEAD	
SNP	**Locus**	**EA[†]**	**EAF[‡]**	**N**	**OR[§] (95CI[‖])**	**P**	**N**	**OR (95CI)**	**P**	**N**	**OR (95CI)**	**P**
rs10116277	9p21	T	0.49	1375/7347	1.10 (1.01,1.20)	2.1e-02	398/6614	1.14 (0.98,1.32)	8.6e-02	1127/8407	1.15 (1.05,1.25)	3.0E-03
rs2383207	9p21	G	0.52	1351/7285	1.10 (1.03,1.17)	6.0e-03	388/6565	1.11 (0.98,1.21)	8.6e-02	1109/8331	1.12 (1.05,1.19)	2.0e-03
rs1333040	9p21	C	0.50	1585/8059	0.89 (0.82,0.96)	4.0e-03	481/7245	0.90 (0.79,1.04)	NS	1263/9314	0.88 (0.81,0.96)	5.0e-03
rs1333049	9p21	G	0.49	1567/7884	1.07 (0.99,1.15)	NS	469/7087	1.11 (0.96,1.27)	NS	1236/9149	1.03 (0.95,1.13)	NS
rs10757274	9p21	G	0.50	1620/8101	1.12 (1.04,1.21)	4.0e-03	488/7263	1.13 (0.99,1.29)	7.3e-02	1292/9387	1.13 (1.04,1.23)	3.0e-03
rs10455872	*LPA*	G	0.24	1711/10683	1.39 (1.21,2.58)	1.1E-06	519/9952	1.08 (0.84,1.39)	NS	1351/12067	1.25 (1.08,1.45)	3.0E-03
rs3798220	*LPA*	G	0.02	1765/11343	1.18 (0.93,1.49)	NS	537/10577	1.09 (0.70,1.72)	NS	1413/12772	1.39 (1.09,1.79)	9.0E-03

[†]Effect allele; [‡]Effect allele frequency; [§]odds ratio;[‖]Confidence interval; All logistic models were correct for age, gender and diabetes status

3.6.3.2 LPA SNPs

In this study we found that the minor allele of rs10455872 was associated with increased risk of CAD and LEAD, and not with IS and that the minor allele of rs3798220 was associated with LEAD but not with CAD or with IS (Table 3-3).

3.6.4 Meta-analysis of rs10455872 and rs3798220 effects in cardiovascular phenotypes

A fixed effect per minor allele of rs3798220 was 1.43 (1.26-1.61), p=5.32E-28 (Figure 3-1) and from the random effects model was 1.43 (1.18-1.72) on risk of CAD. The fixed effect estimate per minor allele for rs10455872 was 1.33 (1.06-1.29), p=1.06E-21 (Figure 3-1) and the random effects estimate was 1.33 (1.19-1.49). Significant heterogeneity was detected in the CAD phenotype so both fixed and random effects are reported.

There was no association of rs107455872 p=0.67 or rs3798220 p=0.30 with IS in the meta-analysis (Figure 3-1). Both LPA SNPs were associated with LEAD: rs107455872 had an OR=1.19 (1.08-1.31) per minor allele, p=4.0E-04 and rs3798220 had an OR=1.36 (1.16-1.63) per minor allele, p=9.0E-04 (Figure 3-1). No significant heterogeneity was detected in the meta-analysis of IS or LEAD allelic effects.

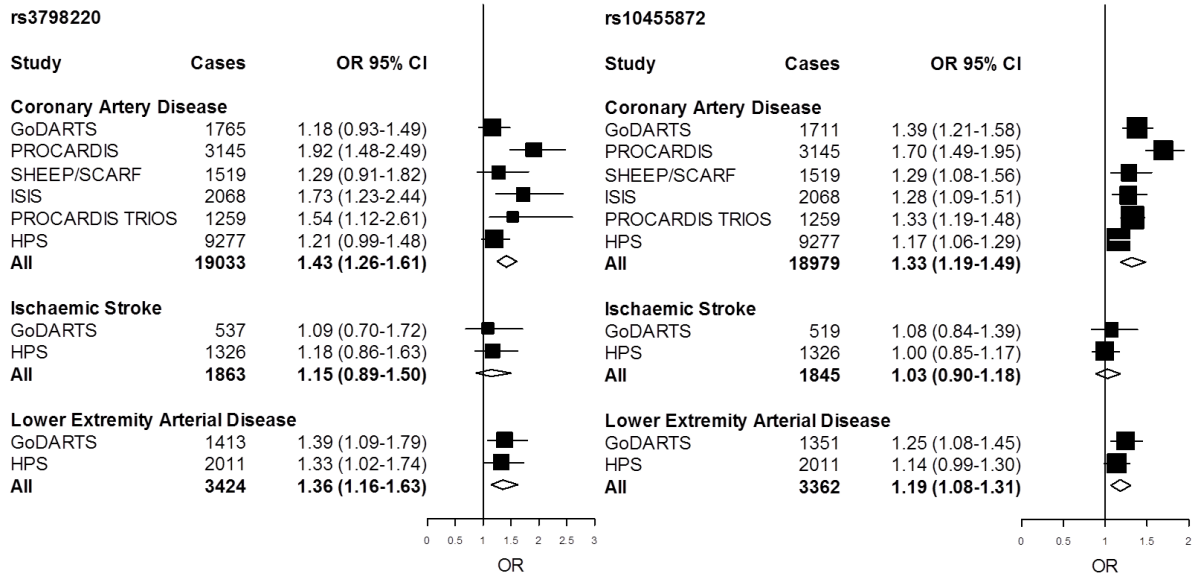

Figure 3-1: Forest plots of the effect of LPA SNPs rs3798220 and rs10455872 on cardiovascular phenotypes combined across studies.

3.6.5 Association between GS for cardiovascular risk factors and their corresponding traits

The GS scores showed a positive correlation and were all significantly associated with their corresponding traits - irrespective of diabetic status (Table 3-4) - and tests for interaction with T2D were not significant.

Table 3-4: Association each genotype score with its corresponding trait in type 2 diabetics and non-diabetics in the GoDARTS study

			Diabetic				Non-diabetic		
Trait*	**SNPs**	**N**	**Beta**	**SE**	**P**	**N**	**Beta**	**SE**	**P**
SBP	29	5583	0.28	0.142	4.47E-02	3902	0.76	0.152	6.050E-07
DBP	29	5583	0.56	0.222	1.16E-02	3902	0.40	0.129	1.77E-03
TCHOL	61	6119	0.88	0.053	1.28E-61	3535	0.98	0.057	4.62E-64
LDL-C	48	4090	0.93	0.042	7.30E-102	2991	0.93	0.048	1.43E-77
HDL-C	52	5838	0.70	0.080	1.65E-18	3489	0.90	0.059	1.66E-52
TG	37	4549	0.74	0.062	2.69E-32	3069	0.40	0.040	9.32E-24
FGLU	17					1886	0.86	0.133	1.29E-10
			OR	**95%CI**			**OR**	**95%CI**	
CAD	45	8964	1.32	(1.17, 1.49)	1.10E-05	4881	1.34	(1.16-1.54)	5.91E-05
T2D†	61	10207	1.16	(1.14, 1.19)	2.79E-86				

*All models were corrected for age and gender; untreated levels of the continuous traits were used for this analysis and the FGLU association was corrected for body mass index; †These scores were tested against a case control design of T2D cases compared to non-diabetic controls where the score has been separated in deciles.

3.6.6 Association of genetic scores for known risk factors and corresponding traits with cardiovascular disease

3.6.6.1 Coronary artery disease

The effects of the GS for SBP, DBP, T2D, LDL-C and FGLU on CAD were consistent with effects observed for measured traits where both GS and trait showed the same direction of effect and were significantly associated with CAD. Even though the GS for serum levels of TG and (directly measured) serum levels of TG per se were both significantly associated with

CAD they had opposing effects on CAD risk. While increasing serum levels of TG were associated with an increased risk of CAD the GS for increasing TG had a protective effect. The GS for total cholesterol (TCHOL) and HDL-C were not substantial modifiers of CAD risk despite the strong association of serum levels of TCHOL and HDL-C with CAD (Table 3-5).

Table 3-5: Association of risk factors and genotype scores for risk factors associated with CAD, N cases = 1441, N controls=7523

	Trait			Genotype score		
Trait	**OR***	**95%CI†**	**P**	**OR**	**95%CI**	**P**
SBP‡,§	1.02	(1.01,1.02)	2.94E-13	1.06	(1.02,1.10)	2.54E-03
DBP	1.03	(1.02,1.04)	6.17E-14	1.10	(1.04,1.17)	7.27E-04
T2D	5.67	(4.8,6.7)	1.01E-91	1.03	(1.01,1.05)	9.85E-03
T1DM				1.00	(0.98,1.03)	NS
TCHOL	1.37	(1.84,2.13)	3.93E-30	1.06	(0.84,1.34)	NS
LDL-C	1.32	(1.3,1.45)	1.44E-09	1.37	(1.14,1.65)	8.46E-04
HDL-C	0.19	(0.16,0.22)	7.19E-106	1.03	(0.64,1.64)	NS
TG	1.14	(1.09,1.18)	4.34E-10	0.82	(0.72,0.94)	3.95E-03
FGLU	1.04	(1.02,1.05)	3.08E-08	2.49	(1.31,4.71)	5.11E-03

*Odds ratio; †Confidence interval; ‡ Systolic blood pressure (SBP), diastolic blood pressure (DBP), T2D (T2D), type 1 diabetes (T1DM), total cholesterol (TCHOL), LDL-C (LDL-C), HDL-C (HDL-C), triglycerides (TG), fasting glucose(FGLU); All models were corrected for age, gender and diabetes status except the association with T2D : SBP and DBP were corrected for hypertensive treatment; TCHOL, LDL-C, HDL-C and TG were corrected for statin treatment.

3.6.6.2 Ischaemic stroke

Direct measures for, TCHOL, LDL-C, HDL-C, TG and FGLU and clinically determined T2D and the GS for CAD, SBP and T2D were associated with increased risk of IS (Table 3-6). The GS for CAD, SBP and T2D showed the same direction of effect as their corresponding physical measures on IS risk. Although the associations were not statistically significant, the TG GS reduced the risk of IS while serum levels of TG increased the risk of stroke, similar to the associations with CAD (Table 3-5 and Table 3-6).

Table 3-6: Association of risk factors and genotype scores for risk factors associated with ischaemic stroke, N cases =426, N controls=6624

	Trait			genotype score		
Trait	**OR***	**95%CI†**	**P**	**OR**	**95%CI**	**P**
CAD				1.05	(1.01,1.09)	1.48E-02
SBP‡,§	1.01	(1.00-1.01)	NS	1.07	(1.10-1.14)	4.70E-02
DBP	1.01	(1.00-1.02)	NS	1.07	(0.97-1.18)	NS
T2D	7.54	(5.20-10.94)	1.90E-26	1.07	(1.03,1.11)	1.24E-03
T1DM				1.00	(0.96,1.06)	NS
TCHOL	1.25	(1.14,1.37)	1.52E-06	0.91	(0.63,1.3)	NS
LDL-C	1.21	(1.05,1.41)	1.10E-02	1.01	(0.73,1.39)	NS
HDL-C	0.18	(0.12,0.27)	7.53E-18	1.25	(0.55,2.84)	NS
TG	1.11	(1.02,1.21)	1.50E-02	0.83	(0.65,1.04)	NS
FGLU	1.04	(1.01,1.07)	4.00E-03	2.66	(0.76,9.26)	NS

*Odds ratio; †Confidence interval; ‡ Systolic blood pressure (SBP), diastolic blood pressure (DBP), T2D (T2D), type 1 diabetes (T1DM),total cholesterol (TCHOL), LDL-C (LDL-C), HDL-C (HDL-C), triglycerides (TG), fasting glucose(FGLU); All models were corrected for age, gender and diabetes status except the association with T2D : SBP and DBP were corrected for hypertensive treatment; TCHOL, LDL-C, HDL-C and TG were corrected for statin treatment.

3.6.6.3 Lower extremity arterial disease

An increase in the physical measurements of known LEAD risk factors TCHOL, LDL-C, FGLU and T2D status were associated with increased risk of LEAD; however, only the GS for FGLU and T2D were associated with increased risk of LEAD and had the same direction of effect as the measured trait. Similar to the relationship with CAD and IS, serum TG and the TG GS were significantly associated with LEAD but while increasing serum levels predicted increasing LEAD risk, a higher GS predicted reduced risk of LEAD. The associations of the CAD GS and T1DM GS were unique to the LEAD phenotype. No association was observed with the GS for CAD but the T1DM GS was significantly predictive of increased LEAD risk (Table 3-7).

Table 3-7: Association of risk factors and genotype scores for risk factors associated with lower extremity arterial disease, N cases =1195 , N controls=8650

	Trait			Genotype score		
Trait	**OR***	**95%CI†**	**P**	**OR**	**95%CI**	**P**
CAD				1.07	(0.96, 1.18)	NS
SBP‡,§	1.00	(1.00,1.00)	NS	1.03	(0.99,1.06)	NS
DBP	0.98	(0.98,0.99)	6.96E-07	1.06	(1,1.12)	3.50E-02
T2D	5.49	(4.55,6.62)	1.73E-70	1.05	(1.02,1.07)	3.93E-05
T1DM				1.04	(1.01,1.07)	1.97E-03
TCHOL	1.15	(1.09,1.21)	9.58E-08	0.85	(0.68,1.05)	NS
LDL-C	1.13	(1.04,1.24)	6.00E-03	0.87	(0.72,1.05)	NS
HDL-C	0.35	(0.29,0.42)	2.56E-26	1.58	(0.95,2.63)	NS
TG	1.07	(1.02,1.11)	2.00E-03	0.83	(0.72,0.95)	9.00E-03
FGLU	1.03	(1.01,1.04)	1.76E-04	2.33	(1.18,4.59)	1.44E-02

*Odds ratio; †Confidence interval; ‡ Systolic blood pressure (SBP), diastolic blood pressure (DBP), T2D (T2D), type 1 diabetes (T1DM), total cholesterol (TCHOL), LDL-C (LDL-C), HDL-C (HDL-C), triglycerides (TG), fasting glucose(FGLU); All models were corrected for age, gender and diabetes status except the association with T2D : SBP and DBP were corrected for hypertensive treatment; TCHOL, LDL-C, HDL-C and TG were corrected for statin treatment.

3.7 Discussion

We have presented algorithms to identify CAD, IS and LEAD cases from the GoDARTS study and validated the IS and LEAD phenotypes using the EMR approach. SNPs in the 9p21 and *LPA* region were associated with EMR derived cardiovascular phenotypes from GoDARTS and exhibited similar effect sizes and direction of effect when compared to published studies. We have shown that while direct measurements of known cardiovascular risk factors are positively associated with risk of cardiovascular disease that the GS for those risk factors do not always reflect that relationship.

The 9p21 and *LPA* loci are reported in GWAS for cardiovascular disease[38, 161, 268-275] and the 9p21 SNPs examined in this study had similar effect sizes across CAD, stroke and LEAD and when compared to published results[270, 272, 276]. Although, the variants investigated in this

study had not been specifically reported for LEAD, variants linked to rs10757274 - rs1333049 and rs10757269 (R^2>0.8, D'>0.92) - had been shown to increase the risk of LEAD and to lower ABI[158, 268]. The *LPA* SNPs were not associated with IS in this study but did show different allelic effects for LEAD and CAD.

SNPs rs10455872 and rs3798220 map to the *LPA* gene and have been associated with lipoprotein (a) levels, an independent risk factor for CAD, IS and LEAD[50, 277]. Our meta-analysis of the published effects of rs10455872 and rs3798220 confirmed the association of these SNPs with CAD and LEAD[50, 162] but did not detect any relationship with ischaemic stroke based on estimates from GoDARTS and the HPS[162]. This may be due to a lack of power to detect effect as a meta-analysis showed a smaller effect of lipoprotein (a) levels on stroke when compared to CAD[278]. The varying effect sizes of *LPA* SNPs on individual cardiovascular disease types compared to the uniform effects observed for 9p21 SNPs indicates that the phenotypes derived from EMR are discrete enough to reveal differences in molecular patho-aetiological pathways. The associations with the GS further supports this premise as some previous associations with GS scores were replicated in this study but the associations with LEAD were novel and distinct from CAD and IS.

A GS that combined the SNP effects of SBP and DBP into one score was shown to increase the risk of CAD and IS[31] and we observed a similar relationship between separate GS for SBP and DBP with CAD and IS. We also confirmed previous findings that the GS for LDL-C increased the risk of CAD by the same magnitude as the serum levels of LDL-C and that there is no association between the GS for HDL-C and CAD, contrary to a marked protective effect of increasing serum HDL-C[260]. Despite the power to detect an association with GS for CAD and LDL-C, these GS were not associated with the LEAD while the GS for T1DM was distinctly associated with LEAD. These associations indicate that LEAD may have different genetic determinants to CAD and IS and different risk factors. Overall, the associations between different GS with CAD, IS and LEAD were usually consistent with the associations observed for physical measures but this was not true for triglycerides.

The GS for TG showed a protective effect from cardiovascular disease while increased to serum levels of TG increased the risk. This is similar to studies related to T2D, a CAD risk factor, where TG raising alleles in *GCKR* have been associated with decreased T2D risk,

improved insulin sensitivity and plasma glucose concentrations[267, 279-281]. The inverse relationship may be explained in part by the mechanism modifying serum TG levels as observed for the I148M variant of *PNPLA3*. *PNPLA3* has been proposed to increase circulating triglycerides by exporting TG particles from the liver into the blood stream and to protect against T2D and insulin resistance in a similar fashion to *GCKR.* Removing TG from the liver prevents lipotoxic liver damage, simultaneously reducing the risk of T2D and insulin resistance[282]. It may be that similar effects are operating in the vasculature and or cardiac tissue with respect to triglyceride efflux and lipotoxicity.

Pleiotropic effects of TG associated SNPs may also explain the inverse relationship between the GS and CAD as only 6 of the SNPs were independent predictors of serum TG and 7 also predicted serum HDL-C and LDL-C levels. Teslovich et al. (2010) found that 13 of the TG associated SNPs included in the GS were also associated with CAD ($p<0.05$) and that the risk allele was not always the TG raising allele[54]. While the mechanisms that increase serum TG levels may reduce CAD risk the effect of the SNPs on other lipid traits may also modify other risk factors and the risk attributed to the TG GS. Presumably more prosaic reasons might include the fact that SNPs strongly related to TG levels may not be the same as variants that predispose to high TG in the context of a metabolic disturbance.

Identifying related cardiovascular phenotypes from a single population allows for the comparison of SNP affects across these phenotypes and provides insights into how SNPs predict disease outcomes both generally and specifically. In the GoDARTS study, the similar effects of 9p21 SNPs on risk of cardiovascular disease indicate that the EMR identification algorithms are identifying individuals with generalised atherosclerotic vascular disease. The associations of the *LPA* SNPs with cardiovascular outcomes was more specific indicating that the algorithms are sensitive to different forms of vascular disease - a premise further supported by the specific associations of the CAD and T1D GS with cardiovascular outcomes. Validating EMR linked biobanks, like GoDARTS, for genetic studies is important for tracing the global clinical impact of disease related loci across correlated phenotypes within a single population. It improves our ability to assess the importance and impact of these loci on multiple disease outcomes, rather than in the isolation of a single phenotype.

Limitations of these analyses include the fact that cardiovascular cases were biased toward acute events that resulted in hospitalisation or death and include individuals with co-morbidities such as T2D. Although individuals for this study were selected from the larger GoDARTS population and were enriched for T2D[225, 226] and were biased toward acute end points it was still possible replicate published genetic associations. Despite a lack of power to detect significant associations with IS effect alleles and effect sizes were consistent with published findings[161] and the identification algorithm is well validated. It is important to note that IS is made up of several stroke sub-types that have different disease pathways and this phenotype heterogeneity may have affected the power to detect significant associations. The ICD codes used to identify CAD from SMR01 have been well validated by the Heart-disease Evidence-based Audit & Research Tayside Scotland (HEARTS) study[283]. Furthermore, associations between established CAD loci and risk of CAD were well replicated in GoDARTS.

This chapter demonstrates how well validated and characterised cardiovascular phenotypes can be derived from EMR data for genomics research and how these phenotypes can be used effectively to investigate the genetics of multiple facets of cardiovascular disease. Importantly this chapter suggests the existence of the shared and specific genetic architecture of LEAD and CAD. The ability to look across a wide range of both clinical and physiological phenotypes will provide insights into the pleotropic relationships between the both risk factors and disease, as well as across the different forms of vascular disease. The CAD algorithm derived in this study was used to contribute summary statistics to a large scale meta-analysis of CAD.

Chapter 4: Large-scale association analysis identifies new risk loci for coronary artery disease

The CARDIoGRAMplusC4D Consortium

This work has resulted in a publication in Nature Genetics, the pdf of which is included as appendix 7. A full list of authors and author contributions are given in appendix 7

Natalie van Zuydam analysed the CardioMetabochip data for CAD and MI in the GoDARTS dataset, submitted summary statistics to the meta-analysis, analysed the GoDARTS data for the subgroup analyses including additional analyses for the manuscript, the gene score analysis for GoDARTS was provided for the analysis. Colin Palmer contributed to writing the manuscript

4.1 Introduction

There are 31 loci that have been identified by large-scale genetic studies for CAD (**Table 1-2**) which explain about 10% of the total heritability[38, 44, 48-52]. The largest study to date included 22,233 CAD cases and 64,762 CAD free controls of European descent from 14 genome-wide association studies[48]. In order to identify additional variants associated with CAD and to explain more of the heritability much larger sample sizes are required but genotyping on genome wide arrays is prohibitively expensive. In order to overcome the cost of genotyping larger sample sets the large consortia designed a custom Illumina array[234].

The CardioMetabochip consists of SNPs for replication and fine-mapped regions of interest contributed by the Body Fat Percentage, CARDIoGRAM (coronary artery disease and myocardial infarction), DIAGRAM [233], GIANT (anthropometric traits), Global Lipids Genetics (lipids), HaemGen (haematological measures), ICBP (blood pressure), MAGIC (glucose and insulin), and QT-IGC (QT interval) GWAS meta-analysis consortia. Loci were selected based on previous findings from GWAS, individual selections from consortia and earlier genetic studies[234]. For the genetics of CAD the purpose of the array was to provide a focused SNP set for replication and to investigate independent effects in fine-mapped regions. The cost of genotyping individuals on the CardioMetabochip is considerable less than genome wide SNP arrays allowing for a large sample size to be genotyped for a set of SNPs and loci likely to be associated with CAD.

SNPs identified as risk loci for CAD from large genetic studies should increase our power to predict future disease based on genetics. Clinically, CAD risk is assessed by combining known risk factors into a score. The Framingham risk score combines clinically measured systolic blood pressure, low density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) and includes age, sex, diabetes and smoking status into a weighted score[284]. Each risk category is assigned a different weight which is added up and compared to a curve to predict the risk of CAD over 10 years as a percentage[284]. A desirable clinical outcome of large scale meta-analyses for CAD would be to combine the loci into a genetic risk score to provide improved prediction of CAD above what is already provided by traditional risk factors.

The main aim of this study was to replicate an LD pruned SNP set of 6222 SNPs which reached nominal significance p<0.01 by Schunkert et al. (2011) in an independent sample of 190,000 genotyped on the CardioMetabochip (CM). The other SNP content on the CM was also investigated for novel loci that may be associated with CAD. External to the meta-analysis we combined the loci identified at an FDR threshold of less than 5% for association with CAD into a CAD gene score. The predictive ability of the CAD genetic risk score was compared to conventional risk factors in an independent group of individuals from GoDARTS.

4.2 Materials and methods

4.2.1 Analyses prepared from GoDARTS

4.2.1.1 Meta-analysis population and genotyping data

Coronary artery disease cases and controls were identified according to the definition under heading 3.2.1. Individuals typed on the CardioMetabochip were included in all analyses submitted for the meta-analysis (Figure 4-1). A full description of the CardioMetabochip genotyping data is given under heading 2.4.5.

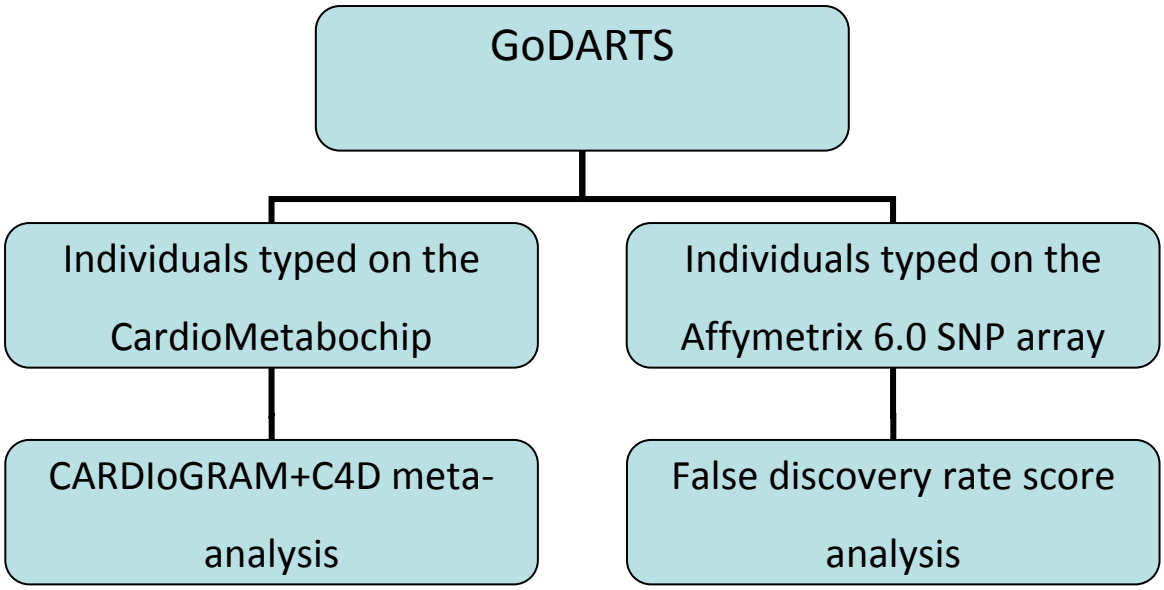

Figure 4-1: A sub-population of GoDARTS was used to estimate summary statistics which were included in the meta-analysis and a separate sub-population was used to calculate an CAD risk score.

4.2.1.2 Statistical methods

Logistic regression analyses were performed in PLINK using prescribed models: all CAD cases with all controls, adjusted for sex and age; male CAD cases with male controls, adjusted for age; female CAD cases with female controls, adjusted for age; early-onset CAD cases with early age of onset (≤50 years) with all CAD controls, adjusted for sex; late-onset CAD cases (>50 years) with all controls, adjusted for sex; and all myocardial infarction cases with all controls, adjusted for age and sex. Age was defined as the recruitment age for controls and the event age for cases.

4.2.2 Large scale replication and meta-analysis

A full description of the meta-analysis methods are given in Appendix 7 under online methods.

4.2.3 CAD genetic risk score

4.2.3.1 Score population

Individuals that conformed to the CAD case and control definition but were not included in the large-meta-analysis were identified in GoDARTS (heading 3.2.1). These individuals were genotyped on the Affymetrix 6.0 array. A full description of these data can be found under heading 2.4.3.

4.2.3.2 SNPs included in the CAD genetic risk score and score calculation

SNPs associated with CAD in the meta-analysis at an FDR of 5% and known CAD SNPs were used to calculate a genotype score. The genotyping data was taken from the Affymetrix 6.0 SNP genotyping array (2.4.3) which had been imputed to HapMap2 and the Wellcome Trust Common Control set typed on the Illumina Human 1M dual using Impute2[238](http://mathgen.stats.ox.ac.uk/impute/impute_v2.html). Where the index SNP was not available a proxy with R^2>=0.8 was used.

4.2.3.2 Derivation of cardiovascular risk predictors from electronic medical records

All cardiovascular risk factors were calculated at the index date which was the time of the CAD event for cases and date of study enrolment for controls. Systolic blood pressure was calculated as a mean of all blood pressure measures before the index date; HDL-C and LDL-C

were also mean values taken before the index date. Smoking status was coded as ever or never with respect to the index date. Age was calculated at the index date.

Cardiovascular risk factors were extracted from multiple EMR sources. Age and sex were extracted from data on demography for individuals in the cohort. Low density lipoprotein cholesterol (LDL-C) and high density lipoprotein cholesterol (HDL-C) were extracted from biochemistry records linked to the cohort (heading 2.1.7). Records of smoking status and regular blood pressure measures are available for the cohort and relevant data was extracted from these sources. Prescribing data was used to identify individuals who were dispensed lipid-lowering and or antihypertensive medication (heading 2.1.8).

4.2.3.3 Genotype score calculation

Genotype scores were calculated for individuals who had not been included in the meta-analysis (Figure 4-1) using the method described under heading 2.9. Risk alleles were weighted using the stage 2 allelic effects estimated in the meta-analysis.

4.2.3.4 Statistical methods

All statistical analyses were conducted using the R statistics software[227]. First the association of the CAD genetic risk score with CAD was established in GoDARTS using a logistic regression analysis corrected for age and sex. Pseudo R square and c statistics were estimated from logistic regression analyses for individual predictors using the lrm function from the rms package. The net reclassification improvement was calculated for all models compared to age and sex and then for all the risk factors compared to all the risk factors and the CAD genetic risk score. These analyses were conducted using PredictABEL. A series of logistic regression models were run to assess the gain in disease prediction inherent in each disease predictor. The basic model included age and sex as predictors and compared the predictive capability of HDL-C, LDL-C, systolic blood pressure, smoking status to the CAD genetic risk score.

4.3 Results

4.3.1 Meta-analysis population from GoDARTS

A total of 619 CAD cases and 2146 CAD free controls were included in the estimation of summary statistics which were submitted to the meta-analysis. The numbers of individuals

included in each of the sub-analyses are given in supplementary Table 1a[285]. 118,658 SNPs were submitted for each analysis and the main analysis had a lambda of 1.03 given in supplementary table 3a[285].

4.3.2 Large scale replication and meta-analysis

A full description of the results is given in Appendix 7. Briefly, the meta-analysis included 63,746 CAD cases and 130,681 CAD free controls. 15 new loci (Appendix 7: table 2) were identified at genome wide significance and 153 loci were associated with CAD at an FDR of 5%. Rs16986953 was genome wide significant in the young CAD and male subgroups and had a significant interaction with age (p=0.033). These loci explain a quarter of the total heritability of CAD attributed to additive genetic variance. Network analysis of the SNPs that passed the FDR threshold showed and enrichment for lipid metabolism signals and to a lesser extent with blood pressure loci and obesity traits.

4.3.3 Score population and characteristics

737 CAD cases and 1771 CAD free controls were included in this CAD genetic risk score analysis. The controls were older than the cases, included more smokers, more statin and antihypertensive treated individuals and had higher HDL-C than the cases (Table 4-1). The cases had higher LDL-C, higher blood pressure and a higher proportion of males (Table 4-1).

Table 4-1: Population characteristics of the coronary artery disease cases and controls used in this study

Trait	**Cases**	**Controls**
Number	737	1771
Age	63.1 (11.2)	65.3 (10.4)
Male %	65	50
Type 2 diabetic %	100	100
History of smoking %	60	72
Low density lipoprotein cholesterol (mmol/L)(SD)	2.3 (0.9)	2.1 (0.8)
High density lipoprotein cholesterol (mmol/L)(SD)	1.2 (0.4)	1.4 (0.4)
Statin treated %	29	63
Systolic blood pressure (mmHg) (SD)	147.3 (18.1)	142.8 (11.9)
Antihypertensive treatment	29	41

4.3.4 The genetic risk score for coronary artery disease and its association with coronary artery disease

153 SNPs identified at an FDR <5% in the meta-analysis were included in the score (Appendix 8): 152 of the SNPs were the actual index SNP reported and rs1034565 was represented by a proxy rs7285377 (R2=D'=1). The CAD genetic risk score ranged from 7.29 to 10.59 for individuals included in the study. The CAD genetic risk score was divided into equal quartiles where each quartile increase in score was associated with a 12% increase in the risk of CAD, the odds ratio was 1.12 (1.04-1.21) and the p value=3.8E-03. When the score quartiles were associated categorically with CAD there was a non-linear increase in CAD risk for increasing score quartiles (Figure 4-2)

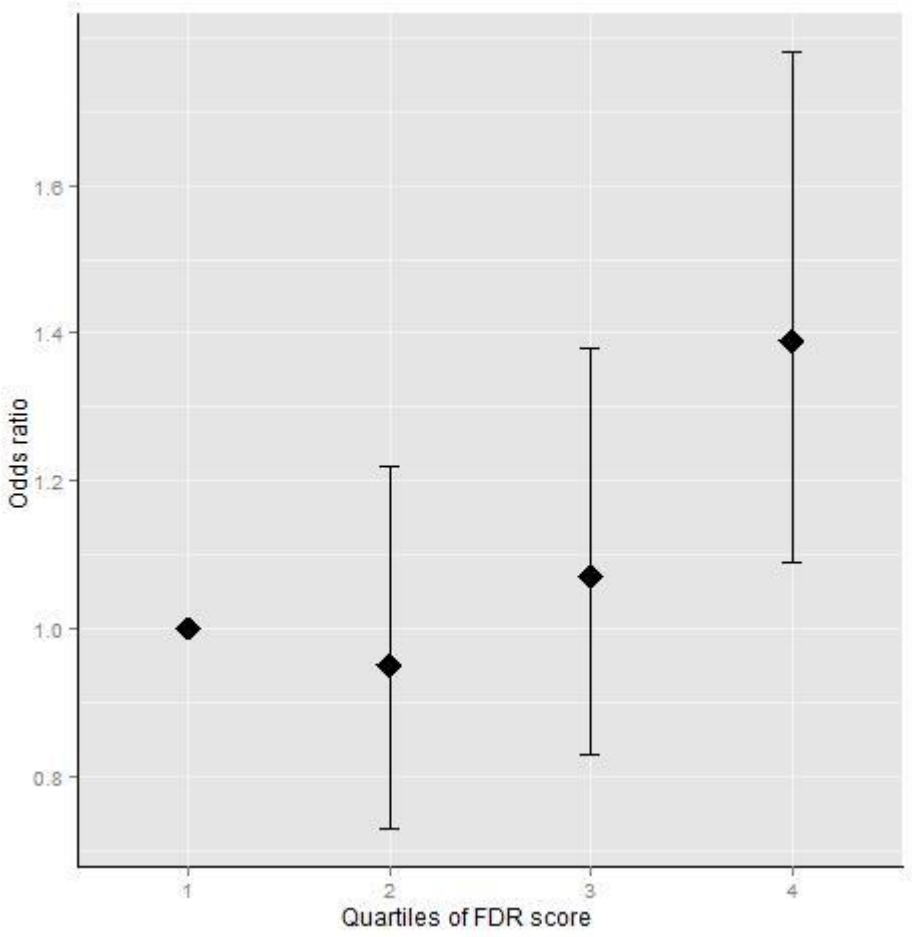

Figure 4-2: A forest plot of the odds ratio of coronary artery disease by quartile of the false discovery rate score.

4.3.5 Evaluation of the CAD genetic risk score as a predictor of coronary artery disease

The pseudo R square, C statistics and net reclassification improvements (NRI) were estimated in a population that was independent of the population used in the meta-analysis. The pseudo R square gives an indication of the ability of a predictor to predict a future outcome and the C statistic is a measure of the area under the receiver operating curve explained by the predictor, and is another measure of how well a predictor classifies individuals as either a case or a control. The net reclassification improvement is also a measure to compare the discriminatory power of two models and is more sensitive than the C statistics in detecting improvements in discriminatory power.

In this study the greatest predictor of CAD outcome was HDL-C which had an R-square =0.244, a C statistic of 0.766 and an NRI of 0.83 [0.75 - 0.92] that was a significantly better predictor ($p<7.8E-03$) when compared to just age and sex alone indicating that it is a good predictor of CAD outcome. The second highest predictor was LDL-C with an R-square value of 0.159, a C statistic of 0.710 and an NRI 0.69 [0.61 - 0.78] that was a significantly better predictor ($p<7.8E-03$) when compared to just age and sex alone. Both these predictors improved the risk prediction above accounting for age and sex. SBP and history of smoking were also good predictors but were not as predictive as the lipid traits (Table **4-2**).

When compared to the other risk factors the CAD genetic risk score was not a good predictor on its own and offered very little additional predictive value above the traditional risk factors which are included in the Framingham equation[284] (Table **4-2**). When the model that included all the risk factors was compared to the model that included all of the risk factors and the CAD genetic risk score the NRI was 0.12 [0.03 - 0.20]and was nominally insignificant p=7.8E-03

Table 4-2: R square and c statistics for individual predictors of coronary artery disease in 737 CAD cases and 1771 CAD free controls

Model	R square	C statistic	NRI [95% CI]
Age and Sex	0.033	0.599	-
Age, sex and history of smoking	0.058	0.626	0.53 [0.44 - 0.61]
Age, sex, systolic blood pressure and antihypertensive treatment	0.051	0.618	0.44 [0.35 - 0.52]
Age, sex, high density lipoprotein cholesterol and statin treatment	0.244	0.766	0.83 [0.75 - 0.92]
Age, sex, low density lipoprotein cholesterol and statin treatment	0.159	0.710	0.69 [0.61 - 0.78]
Age, sex and all risk factors	0.218	0.753	0.80 [0.71 - 0.88]
Age, sex and CAD genetic risk score	0.041	0.606	0.12 [0.03 - 0.20]*
Age, sex, all risk factors and the CAD genetic risk score	0.224	0.756	0.82 [0.73 - 0.90]

*Nominally insignificant after multiple testing p=7.8E-03

4.4 Discussion

The largest meta-analysis for CAD revealed 15 new loci and enrichment for loci in lipid pathways. Rs16986953 was also found to interact with age in the subgroup analyses. Bringing the total number of confirmed CAD associated loci to 46. 153 loci associated with CAD at an FDR of 5% explain 10.6% of the narrow sense heritability. The predictive power of the FDR loci for CAD was not assessed in the meta-analysis paper but we investigated this aspect in a group of GoDARTS individuals that were independent of the meta-analysis population.

In this chapter we found that a gene score made up of 153 independent loci for CAD was not as strong a predictor of CAD outcome as serum measures of LDL-C and HDL-C. When the score was included with other predictors it did not notably increase the prediction of CAD above the information provided by conventional risk factors which was in agreement with results that have not been published for the Atherosclerosis Risk in Communities (ARIC) study. This would indicate that we are still unable to use the current set of loci to predict CAD over and above traditional risk factors.

The CAD genetic risk score does not include variants with a minor allele frequency (MAF) of less than 2% and excludes SNPs like rs3798220 and rs10455872 from the *LPA* locus which

have a large effect on CAD risk[50]. It may be that other predictors of CAD are variants that have a MAF of less than 1% and haven't been detected by conventional GWAS and meta-analysis methods. Large-scale analyses of genotyping data imputed to 1000G are currently being undertaken in order to identify rare variants that influence the risk of CAD. These efforts may explain a proportion of the missing heritability for CAD and increase our power to use genetics to predict CAD. Constructing a risk score for CAD may not be as straightforward as a weighted score based on the variants influence on CAD.

Many of the loci detected at the FDR threshold are linked to the lipid metabolism pathway. In this study we found that LDL-C and HDL-C are the strongest predictors of CAD but the risk score did not improve the predictive power beyond these measured risk factors. We know that lipid raising alleles do not always confer increased CAD risk and many have pleiotropic effects on multiple lipid subgroups[54](Chapter 3). It may be more useful to use a more comprehensive score which combines known loci for CAD with loci for known risk factors like nicotine dependence and hyperlipidaemia. There are many pitfalls in this approach and large Mendelian randomisation studies would have to be undertaken to identify the best genetic instruments to include in a CAD genetic risk score. Working from this basis the genetic loci included in the score and the architecture of the score would have to be refined in large prospective studies.

A weakness of these analyses is that the sample size used to investigate the effects of the CAD genetic risk score and known risk factors was small, although our findings agree with those obtained by ARIC. Given that the risk factors included in this study are strong predictors and the genotype score is a well powered instrument when compared to individual genotypes we were still powered to detect an effect. The individuals included in the analysis were all diabetic and the allelic effects were estimated from a mixture of diabetic and non-diabetic individuals. The strengths of our analyses are that we used a completely independent sample from the population used to estimate the allelic effects of SNPs on CAD risk. The individuals in GoDARTS are well characterised and the values given for each of the risk factors were estimated from multiple measurements lending strength to the statistical models used in this chapter.

30% of the heritability of CAD remains unexplained and further studies are required to identify variants that contribute to the missing heritability. Despite the 3-4 fold increased risk of CAD in individuals with T2D, large-scale association studies have not identified an overlap in loci that predict T2D and also predict CAD. Some of the missing heritability may be found in these loci. To follow on from the large meta-analysis we will conduct a meta-analysis of CAD in T2D to identify novel loci associated with CAD in T2D individuals and to assess the role of known loci in determining CAD in T2D populations.

Chapter 5: A meta-analysis of Coronary Artery disease in 18,158 patients with type 2 diabetes

Author contributions:

Natalie van Zuydam (University of Dundee) – presented the analysis to CARDIoGRAM+C4D executive and steering committees for approval; collected the data from CARDIoGRAM+C4D member and analysed the GoDARTS data; wrote the analysis plan and meta-analysis plan for the meta-analysis; performed the meta-analysis and annotated the results

Claes Ladenvall (University of Lund) – collected summary statistics from ENGAGE partners; prepared the analysis for SDR and DGI; performed the meta-analysis in parallel with the Dundee centre

Summary statistics were prepared by analysts from: Cleveland Clinic GeneBank Study; COROGENE study; Duke Cardiac CATHeterization GENetics (CATHGEN) study; Estonian Genome Center, University of Tartu; The Genetics of Diabetes Audit and Research in Tayside Scotland; Lübeck Registry of Structural Heart Disease /KORA (Kooperative Gesundheitsforschung Ludwigshafen Risk and Cardiovascular Health Study; Scannia Diabetes Register; Premature Atherosclerosis Study; Pfizer-MGH-Broad Monica Risk Genetics Archiving and Monograph; The Hellenic Study of Interactions between SNPs and Eating in Atherosclerosis; Susceptibility Dietary, life style, and genetic determinants of obesity and metabolic syndrome

Considerable intellectual advice was given by members of the Work Package 1 analysis team for SUMMIT: Nigel W Rayner, Rona Strawbridge, Harshal Deshmukh, Ninna Sandolm and Emma Alquivist

Prof Colin N A Palmer was responsible for conceiving and writing the project proposal for this study that was submitted to the CARDIoGRAM+C4D steering and executive committees and was involved with procuring data

The SUMMIT proportion of this study was supported by: Mark McCarthy, Helen Colhoun and Leif Groop

5.1 Introduction

In the previous chapter we presented the results from a large-scale meta-analysis of coronary artery disease (CAD) in mixed diabetic and non-diabetic populations. This work identified 46 loci that were associated with increased risk of CAD[38, 48, 285]. The estimated variance explained by 153 loci which passed a false discovery rate threshold explained 10.6% of the additive variance which is just a quarter of the heritability estimated from twin and family studies[31] and is not much larger than the component explained by the previously established loci[48]. We know that the main pathways that have been implicated in the pathology of CAD are inflammation and lipid metabolism pathways (Chapter 4). However, diabetes increases the risk of CAD 3-4 fold[30] and is the most common macrovascular complication of diabetes. Despite evidence of an epidemiological association between CAD and T2D the largest meta-analyses for CAD and T2D, which included GoDARTS, found no overlapping risk loci at genome wide significance[266, 285].

Heritability estimates for subclinical markers of cardiovascular disease in T2D populations are estimated between 30-50%, indicating that there is a large heritable component[286]. Based on the epidemiological relationship and the heritability component it is interesting that so far no overlap in risk loci has been observed for these two diseases in conventional meta-analyses. This may be due to the lack of analyses specifically designed to identify an overlap between CAD and T2D in large genetic studies.

In this chapter we aimed to identify loci associated with CAD specifically in T2D populations. We aimed to identify novel loci for CAD in T2D populations by combining summary statistics from the European Network for Genetic and Genomic Epidemiology (ENGAGE), Surrogate markers for micro- and macro-vascular hard endpoints for innovative diabetes tools (SUMMIT) and CARDIoGRAM+C4D in a meta-analysis. We also examined the role of known CAD and T2D loci in the same cohorts.

5.2 Materials and Methods

5.2.1 Study Design Summary

We combined summary statistics estimated from patients with type 2 diabetes, from 14 genome wide association studies (GWAS) that included 5142 CAD cases and 6828 CAD free controls with 8 CardioMetabochip (CM) studies that included 1813 cases and 4375 controls, in a meta-analysis.

5.2.2 Phenotype

5.2.2.1 Type 2 diabetes

Type 2 diabetic individuals were identified by individual studies according to a clinical diagnosis of T2D based on either the world health organisation (WHO) or American diabetes association (ADA) guidelines. If the type of diabetes was unknown we assumed type 2.

5.2.2.2 Coronary artery disease cases and controls

Individuals with coronary artery disease were identified by individual studies based on a number of criteria given in Table 5-1. Controls were identified to be free of CAD by individual studies. Since the diagnosis of T2D can vary with respect to the diagnosis of CAD in the clinical setting, CAD cases with a diagnosis of T2D up to 5 years after the event were accepted as T2D CAD cases and the same criteria applied to CAD free controls who were classified as T2D if they had a date of diagnosis up to 5 years after the date of study enrolment.

5.2.3 Genotypes and imputation

Operational details of the 14 GWAS studies are given in Table 5-3 and Table 5-4, and the 8 CM studies are given in table 5-5. The GWAS studies were typed on the Affymetrix 6.0, Illumina Human Omni1-Quad and the Illumina Omni Express SNP arrays (Table 5-3). To obtain an increased marker set that was comparable across all GWAS platforms individual studies imputed their GWAS data to the HapMap2 CEU reference panel. Individual studies applied quality control pre-imputation and the details of that QC are given in Table 5-3. Either MACH or IMPUTE2 were used to impute missing genotypes yielding a maximum of 2,619,962 SNPs (Table 5-4).

5.2.4 Statistical methods

5.2.4.1Individual studies

The details of the analyses performed and the software used by individual studies are given in Table 5-4. Briefly analysis software that took genotype uncertainty into account were used to analyse imputed data. Imputed SNPs were tested for their association with CAD using a log-additive model frequentist tests adjusted for age (onset of the first event for cases or time of recruitment for controls), gender and centre specific covariates to account for population structure was applied to imputed SNPs. Directly typed SNPs were modelled log-additively in a logistic regression that was corrected for age, gender and centre specific covariates to account for population structure.

5.2.4.2Data cleaning steps applied before conducting the meta-analysis

Quality control of the data was performed centrally. SNPs with a MAF lower than 1% or those with less than 10 minor allele homozygotes calculated as (2 x (Number of cases) x MAF)) were removed from individual studies. Specific quality control criteria were applied to directly typed and imputed SNPs separately. Directly typed SNPs with a MAF greater than 5% were removed if the p value for the exact test for Hardy-Weinberg equilibrium (HWE) was less than 5.7E-07 and for SNPs with a MAF less than 1% with a less than 1E-04. Imputed SNPs were removed if they had an R^2 value less than 0.3 if imputed with MACH and an R^2 value less than 0.4 if imputed using IMPUTE2. If imputed SNPs were analysed using SNPTEST then imputed SNPs were removed if the proper info value was less than 0.4. Pairwise allele frequency plots were drawn for each study for imputed and directly typed data separately to check for allele frequency outliers and strand differences amongst the summary meta-data submitted for analysis.

5.2.4.3Statistical model checking

The summary statistics for individual studies were checked by drawing plots, for directly typed and imputed SNPs separately, of the beta distribution by MAF category (0-0.01, 0.01-0.05, 0.05-0.2, and 0.2-0.5). A deviation of the mean beta values from zero and scattered outliers were an indication of problems with the models used or low quality SNPs. To check for population stratification quantile-quantile plots were drawn from the p values submitted by each study and lambda was calculated.

5.2.4.4Fixed effects meta-analysis

The primary analysis comprised of two stages: a meta-analysis of the GWAS and a combined analysis of the GWAS cohorts and the CM cohorts. Our default meta-analysis used a fixed-effect model with inverse variance weighting and a calculation of two homogeneity statistics: Cochran's *Q* and I^2 using GWAMA (http://www.well.ox.ac.uk/gwama/)[252]. A Full description of these methods and statistics are given under heading 2.11. Individual studies were corrected for genomic inflation using GWAMA and the overall meta-analysis was corrected for genomic inflation by adjusting the standard error of the estimates by the inflation factor (SE*sqrt(λ)) and recalculating the p values. When there was no indication for heterogeneity for a SNP based on a Cochran's Q p value greater than 0.01 the fixed effect model was maintained.

5.2.4.5Random effects meta-analysis

When heterogeneity was present (Cochran's Q p value <0.01) the study with the most extreme result was excluded and the meta-analysis was repeated. If heterogeneity was still apparent, we adopted and reported a random-effects model[246] for that SNP excluding the outliers. The random effects model was applied using MetaSoft (http://genetics.cs.ucla.edu/meta/)[246]. A full description of the random effects model is given under heading 2.11.3.

5.2.4.6Comparison of effects estimated form the meta-analysis with known coronary artery disease loci and Type 2 diabetes loci

The allelic effects estimated from this study were compared with published effects for 46 CAD and 65 T2D loci. The effect estimates for the CAD loci were taken from the largest meta-analysis to date, and is described in chapter 4[285]. The T2D loci were taken from the largest meta-analysis of type 2 diabetes[287]. Within study effect estimates were plotted against published estimates using the plot function in R[227].

Power calculations were performed using the PS – Power and sample size calculator (http://biostat.mc.vanderbilt.edu/PowerSampleSize)[288] to identify loci that we should be able to detect given the size of the study and a nominal significance level of 0.05. The within study estimates were compared with published estimates for loci which we had the power

to detect and for those that reached nominal significance of p<0.05 by testing for heterogeneity. The heterogeneity calculation is given under heading 2.12.

5.2.4.7Testing for independent effects in the 9p21 region

CAD and T2D share a locus 9p21, rs1333049 is the lead SNP for CAD and rs944801 (represented by rs7030641, R^2=0.966, D'=1) and rs10811661 are independent signals for CAD in the 9p21 region and the SNPs are not in strong linkage disequilibrium. To determine if the signals are independently predictive of CAD in T2D, we investigated their relationship with CAD in a sub-population of GoDARTS that had been genotyped on the CardioMetabochip. All analyses were conducted in R[227] using the glm function from the stats package. Binary logistic regression analyses were corrected for age, sex, and for diabetes status where necessary.

First the SNPs were associated with CAD in a mixed population and T2D to determine if the SNPs were predictive of CAD and T2D in GoDARTS. The SNPs were then tested for interaction with diabetes status with CAD as an outcome. The second analysis restricted the association of the SNPs with CAD to just individuals with T2D to ascertain their affect and association with CAD in a T2D background. Finally all the SNPs were included in the same CAD model in T2D individuals to identify any changes in the effect estimates of the SNPs by including the other in the model.

5.3 Results

5.3.1 Study population

The case phenotype definitions largely overlapped to include at least one of the following criteria: myocardial infarction, unstable angina or corrective procedures (Table 5-1). Cases were identified either by angiography or from hospital admission and death records (Table 5-1). All controls were identified as CAD free individuals and excluded the case definition criteria (Table 5-1).

Age in cases ranged from 41-66 years in cases and 40-72 years in controls; the current smoking status of in cases ranged from 15-85% and controls between 15 and 88% and the BMI in cases ranged from 28 to 34 kg/m^2 in cases and controls 28-33 kg/m^2 (Table 5-2).

Table 5-1: Population characteristics for all cohorts in the CAD meta-analysis

Study	Full name	Reference	Ethnicity	Region of Recruitment	Phenotype	Case	Control
Cleveland Clinic GeneBank Study	Cleveland Clinic GeneBank Study	(21475195 [289])	European	Cleveland, OH	CAD	Subjects undergoing elective diagnostic cardiac catheterization procedure	Controls free of CAD at baseline
Corogene	COROGENE study	[290]	European	Finland	MI	Hospital admission codes and death records for: MI only	Controls free of CAD
DGI	Data not yet submitted						
CATHGEN	Duke Cardiac CATHeterization GENetics (CATHGEN) study	www.cathgen.duhs.duke.edu	Caucasian	USA	CAD	at least one of these criteria: significant CAD on cath, ever MI (history, current or subsequent), hx PCI, hx CABG	no CAD on cath, no hx of cabg or PCI, no ever MI
EGCUT	Estonian Genome Center, University of Tartu	[291]	European	Estonia	CAD	Hospital admission codes	Controls free of CAD
Go-DARTS	The Genetics of Diabetes Audit and Research in Tayside Scotland	[225]	European	Scotland	CAD	Hospital admission codes and death records for: MI/Unstable Angina/Corrective procedures (PTCA &	Controls free of CAD

Study	Full name	Reference	Ethnicity	Region of Recruitment	Phenotype	Case	Control
						CABG)	
KORAF3	Lübeck Registry of Structural Heart Disease /KORA (Kooperative Gesundheitsforschung	[48] [292]	European	German	CAD	Consecutive patients referred for coronary angiography, classified as CAD/MI cases based on the coronary angiogram; CAD < 65 y in males, CAD < 70 y in females	Controls free of CAD
LURIC	Ludwigshafen Risk and Cardiovascular Health Study	[42]	European	Germany	CAD	Angiography	Controls free of CAD
PennCath	Data not yet submitted						
MedStar	Data not yet submitted						
MIGen	Data not yet submitted						
RS	Data not yet submitted						
SDR	Scannia Diabetes Register						Controls free of CAD

Study	Full name	Reference	Ethnicity	Region of Recruitment	Phenotype	Case	Control
EGCUT	Estonian Genome Center, University of Tartu	[291]	European	Estonia	CAD	Hospital admission codes	Controls free of CAD
IMPROVE		[293]	European	Finland, Sweden, Netherland, France, Italy	CAD		
SCARFSHEEP		[294]	European	Sweden	MI/CAD	First confirmed myocardial infarction	No history, symptoms or signs of cardiovascular disease
AMC-PAS	Premature Atherosclerosis Study		European	Netherlands	CAD	MI, surgical or percutaneous revascularisation, coronary angiograph with ≥70% stenosis in a major epicardial artery	
PMB	Pfizer-MGH-Broad		European	Finland, Sweden	CAD/MI		Controls free of CAD
MORGAM	Monica Risk Genetics Archiving and Monograph	[295]	European	Finland	CAD	Hospital admission codes and death records for: MI/Unstable Angina/Corrective procedures (PTCA & CABG)	
THISEAS	The Hellenic Study	[296]	European	Greece	CAD/MI	First-ever mi before	<30% stenosis

Study	Full name	Reference	Ethnicity	Region of Recruitment	Phenotype	Case	Control
	of Interactions between Snps and Eating in Atherosclerosis Susceptibility					age of 70 yrs.; first CAD (>50% stenosis in one of the three main coronary vessels assessed by coronary artery angiography - no history of ACS)	assessed by coronary artery angiography or negative stress test; age matched without MI/CAD history
DILGOM	Dietary, life style, and genetic determinants of obesity and metabolic syndrome	[297]	European	Finland	CAD	All CAD: Incident definite or possible MI or coronary death, or unstable angina during follow-up, Coronary revascularization during follow-up, Documented MI at baseline, or an unclassifiable coronary death during Follow-up. MI: Definite myocardial Infarction.	Non-cases from the same population-based longitudinal cohort study

Table 5-2: Population characteristics for all studies included in the meta-analysis of coronary artery disease in type 2 diabetic individuals

Study	**Mean age (SD) years cases/controls**	**N Diabetic CAD/MI cases/controls**	**% current smoker CAD/MI cases/controls**	**BMI (kg/m^2) CAD/MI cases/controls**
CCF	61.54 (11.06)/ 72.79 (6.42)	510/30	84.7%/15.3%	31.02 (6.38)/ 29.46 (5.67)
Corogene	66.3423 (11.68)/ 55.58 (12.19)	257/117	38.1%/24%	
DGI	67.26/62.11	365/524	39.7%/35.7%	28.25 (4.05)/ 28.54(4.55)
CATHGEN	63.51 (11.27)/ 58.47 (11.79)	397/132	58.1%/41.3%	32.13 (7.13)/ 35.21 (9.68)
EGCUT	65.03(12.76)/ 49.84(19.67)	171/214	14.42%/30.5%	31.18(5.94)/ 31.04(6.04)
GoDARTS	62.75 (10.97)/ 65.43(10.07)	877/2187	74.8%/61.2%	30.72 (5.45)/ 31.1 (5.74)
KORA F3	67.57 (6.53)/ 66.44 (7.60)	51/136	23.53%/11.03%	30.96 (5.18) /31.08 (4.51)
LURIC	59.36 (10.56) / 60.13 (11.53)	934/283	70.2%/51.1%	28.06 (4.01)/ 28.75 (4.64)
PennCath	Data not yet submitted			
MedStar	Data not yet submitted			
MIGen	Data not yet submitted			
RS	70.5 (7.60)/ 68.80 (8.30)	188/380	24.5%/25.9%	27.00(3.50)/ 27.80 (4.00)

Study	Mean age (SD) years cases/controls	N Diabetic CAD/MI cases/controls	% current smoker CAD/MI cases/controls	BMI (kg/m^2) CAD/MI cases/controls
SDR	Data not yet submitted			
EGCUT	59.64(10.53)/55.7 (10.9)	223/517	20.6%/25.4%	33.65(5.30)/ 33.33(5.36)
IMPROVE	65.73 (5.98)/ 64.15 (5.38)	47/907	19.4%/14.6%	30.15 (5.11)/ 29.19 (4.63)
SCARFSHEEP	57.62 (7.33)/ 58.50 (7.20)	237/104	40.7% /27.1%	28.34 (4.44)/ 27.95 (4.43)
AMC-PAS	41.34(5.76)/ 40.11(5.99)	62/6	73.6%/88.5%	29.66(4.89)/ 30.28(4.38)
PMB				
MORGAM	59.3 (8.65)/ 55.73 (10.52)	195/281	28.13%/27.38%	--
THISEAS				
DILGOM	63.79 (7.68)/ 51.82 (13.50)	39/267	18.65%/17.4%	--

5.3.2 Imputation quality control and analysis methods for individual studies

Each cohort was genotyped separately on a number of different platforms. In order to make the GWAS studies more comparable genotypes were imputed from HapMap2. Pre-imputation quality control and imputation to HapMap2 were conducted by individual studies. The pre-imputation quality control criteria and the number of directly typed SNPs included by each study are given in Table 5-3, generally different thresholds for missingness by individual and by genotype were applied. Hardy-Weinberg equilibrium p value thresholds were also imposed. Table 5-4 includes the details of the imputation software and the reference panel used to impute genotypes, generally either MACH or IMPUTE2 were used to impute genotypes from HapMap2. Directly typed and imputed genotypes were analysed separately by individual cohorts and the analysis software used to generate the summary

statistics as well as the number of directly typed and imputed SNPs submitted are given in Table 5-4.

The CardioMetabochip is a much smaller array which contains very different content to the genome wide arrays. This array contained content from a number of different genome wide array platforms so they were not imputed. The software used to analyse the CardioMetabochip data by individual cohorts as well as the model run and the number of SNPs that were submitted with summary data are given in table 5-5.

Table 5-3: Genotyping platforms and pre-imputation QC for genome wide studies

Study	Genotyping Centre	Genotyping Array	Calling algorithm	Pre-Imputation QC Exclusion Criteria	Sample call rate	SNP call rate	HWE
CCF	University of Ottawa	Affymetrix 6.0	Birdseed	Filtered samples for IBS, heterozygosity, sex consistency and standardized IBS eigenvector outliers	97%	95%	1.00E-04
Corogene	Sanger	Illumina 610K	Illuminus	failed gender check, Low genotype frequency(<95%), excess heterozygosity, non-European, excess relatedness, low MAF,	>97%	95%	1.00E-06
DGI	Broad Institute	Affymetrix 500K	Birdseed	MAF>0.01; Multiple hits;; Sex check; Cryptic relationships	>95%	>95%	1.00E-06
CATHGEN	Duke	Illumina HumanOmni1-Quad_v1_0_C	GenomeStudio Genotyping module	MAF < 1%	99%	99%	None
EGCUT	Estonian Biocenter, Genotyping Core Facility	OmniExpress	GenomeStudio	Sex mismatch; cryptic relatedness	95%	95%, MAF 0.01	1.00E-06
Go-DARTS	Sanger	Affymetrix 6.0	CHIAMO		99%	99% if MAF<0.05 & 0.95 if	1.00E-06

Study	Genotyping Centre	Genotyping Array	Calling algorithm	Pre-Imputation QC Exclusion Criteria	Sample call rate	SNP call rate	HWE
						MAF>0.05	
KORAF3		Affymetrix 500K		Call rate 95%; Sex check; Cryptic relationships	95%	95%	1.00E-06
LURIC	LURIC Study non-profit LLC	Affymetrix 6.0	Birdseed v2		99%	98%	1.00E-04
PennCath	Data not yet submitted						
MedStar	Data not yet submitted						
MIGen	Data not yet submitted						
RS		Illumina 550K		Sex check; IBS check; Heterozygosity check; Ethnic outliers	97.5%	90%	1.00E-06
SDR	Data not yet submitted						

Table 5-4: Imputation procedures and reference panels for genome wide cohorts

Study	Imputation Software	Reference Panel	Analysis software	Analysis Model	Total SNPs included in analysis diabetic only [genotyped/imputed]
CCF	MACH 1.0.16	Hap map R22	PLINK	Logistic regression	2543887
Corogene	MACH1.16	Hap Map 2 release 22 CEU	ProbABEL	Logistic regression	2543887 [521288/2022599]
DGI	MACH	HapMap2	PLINK/ProbABEL	Logistic regression	
CATHGEN	Impute2	HapMap2/ Illumina Duo	SNPTEST	Logistic regression	800488/2647475
EGCUT	Impute2	HapMap2	SNPTEST2	Logistic regression	615575/2523018
Go-DARTS	Impute2	HapMap2/ Illumina Duo	SNPTEST	Logistic regression	2500000
KORAF3	Impute2	Hap Map 2	SNPTEST	Logistic regression	
LURIC	MACH	HapMap2	QUICKTEST /PLINK	Logistic regression	2543887
PennCath	Data not yet submitted				
MedStar	Data not yet submitted				
MIGen	Data not yet submitted				
RS	MACH	HapMap2	ProbABEL	Logistic regression	
SDR	Data not yet submitted				

Table 5-5: Genotyping information and analysis model applied for the CardioMetabochip studies

Study	Genotyping centre	Genotyping array	Calling algorithm	Analysis software	Analysis model	Total SNPs included in analysis diabetic only
EGCUT	Estonian Biocenter, Genotyping Core Facility	Illumina CardioMetabochip	GeneCall	PLINK/SNPTEST2	Logistic regression	129838
IMPROVE	Uppsala, Sweden	Illumina CardioMetabochip	GeneCall	PLINK	Logistic regression	185704
SCARFSHEEP	Uppsala, Sweden	Illumina CardioMetabochip	GeneCall	PLINK	Logistic regression	185704
AMC-PAS	SANGER	cardiochip IBC Array	Beadstudio	PLINK	Logistic regression	48105
PMB	Broad institute	Illumina CardioMetabochip	Birdseed	PLINK	Logistic Regression	121812
THISEAS	Sanger	Illumina CardioMetabochip	GenoSNP	PLINK	Logistic regression	120383
MORGAM	Sanger	Illumina	GenCall	PLINK	Logistic	196610

Study	Genotyping centre	Genotyping array	Calling algorithm	Analysis software	Analysis model	Total SNPs included in analysis diabetic only
DILGOM	Helsinki	CardioMetabochip Illumina CardioMetabochip	Illuminus	PLINK	regression Logistic regression	183812

5.3.3 Data cleaning and statistical model checking

Three different types of plots were drawn to check the integrity and the accuracy of the summary statistics submitted by individual cohorts. Pairwise allele frequency plots were drawn for genotyped and imputed data (Figure 5-1A), which should produce a diagonal line through 0:0 and 1:1. Deviations from that line may indicate either allele frequency differences or stand inconsistencies as can be seen in Figure 5-1A. Effect coefficients plotted by minor allele frequency category indicate whether there are any inconsistencies by minor allele frequency or a general problem with the statistical model. We would expect to see a normal distribution of the beta coefficients around a mean of 0. Deviations from that mean line as seen in Figure 5-1B indicates that there are some summary statistics for SNPs with a minor allele frequency less than 5% which are not normally distributed. Effect estimates like this need to be examined carefully to identify the source of the bias. Similarly quantile to quantile plots are used to determine if there is any deviation in the p values from the expected values. We expect some deviation in the tail of the plot if there are variants that are associated with CAD but a systematic inflation or a deflation as see for PennCath in Figure 5-1D reflected in a lambda of 0.94 needs to be scrutinised. In this case it is due to a depowered analysis that contains too few CAD cases but inflation can also be indicative of population structure.

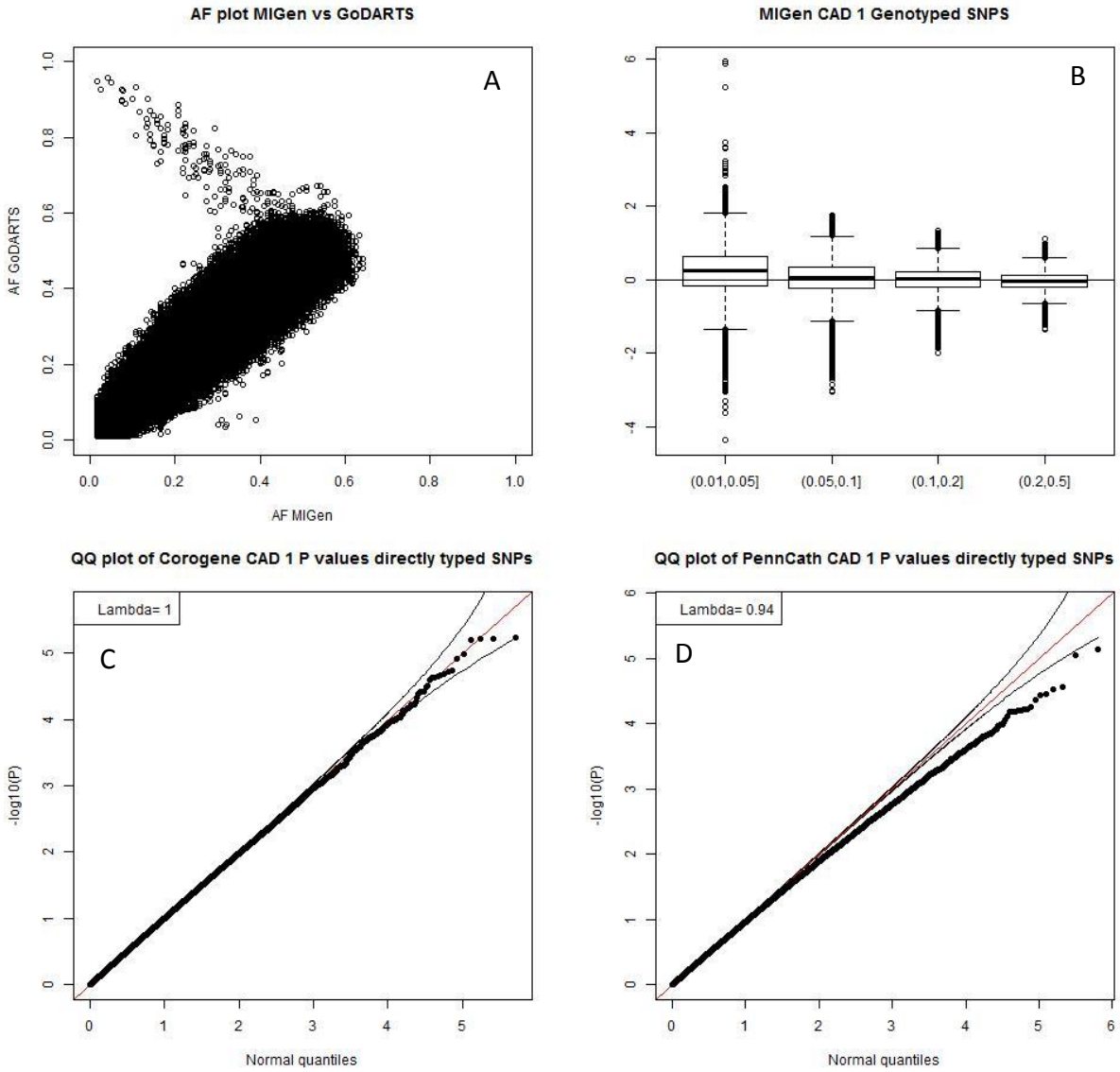

Figure 5-1: Data cleaning and model checking plots are drawn from summary statistics prior to meta-analysis. A – A pairwise allele frequency plot. B- A box plot of beta coefficients (ln(OR)) by minor allele frequency category; C and D - Two quantile- quantile plots from p values submitted by Corogene and PennCath

5.3.4 Post meta-analysis QC

Based on the data checking plots drawn for each cohort quality control criteria were applied to summary statistics from each cohort. All strand issues were rectified to ensure that the combined summary statistics were combined correctly by effect allele. The number of genotyped and imputed SNPs included in the meta-analysis for the GWAS studies and CM studies are given in Table 5-6. Briefly 8414213 imputed and genotyped SNPs were included

in the combined meta-analysis. After applying quality control criteria 2295146 SNPs were analysed using the fixed effects method and 17904 were analysed using random effects.

Table 5-6: Description of the number of SNPs that were included in the meta-analysis and the genomic inflation factors of each cohort.

Cohort	**Platform**	**# SNPs**	**# directly typed SNPs post QC**	**# imputed SNPs post QC**	**Lambda directly typed**	**Lambda imputed**
CCF	GWAS	2451062	-	2357360	-	1.02
Corogene	GWAS	2543888	513724	1816006	1.00	1.00
DGI	GWAS	2398087	383527	1937384	1.00	1.01
DUKE	GWAS	3447964	724138	1895824	1.00	0.99
EGCUT	GWAS	2523020	574838	1716794	1.02	1.01
Go-DARTS	GWAS	2632475	711373	1885991	1.01	1.01
Go-DARTS-SUMMIT	GWAS	2332500	578594	1552188	1.00	1.01
KORA F3	GWAS	2635496	275465	1543479	1.06	1.06
LURIC	GWAS	2543888	607266	1762568	0.99	1.00
MedStar	GWAS	5399009	655592	4743417	0.95	1.01
MIGen	GWAS	727228	686824	-	0.99	-
PennCath	GWAS	5242195	626345	4615850	0.97	0.98
RS	GWAS	3105354	506650	1811761	0.98	0.98
SDR	GWAS	2356360	613049	1714553	0.99	0.99
GWAS total			**8414213**	**2295146**	**1.02**	

Cohort	Platform	# SNPs	# directly typed SNPs post QC	# imputed SNPs post QC	Lambda directly typed	Lambda imputed
EGCUT	Cardio Metabochip	129839	112901		1.07	
IMPROVE	Cardio Metabochip	184733	57890		0.96	-
PAS	IBC	39799	23585		1.20	-
PMB	Cardio Metabochip	130735	120218		1.01	-
SCARF/SHEEP	Cardio Metabochip	184734	83127		1.06	-
MORGAM	Cardio Metabochip	196610	102843		1.02	-
THISEAS	Cardio Metabochip	196725	102322		1.04	-
DILGOM	Cardio Metabochip	183812	55900		0.96	-
CardioMetabochip total		**196725**	**149871**		**1.01**	

5.3.5 Fixed effects meta-analysis

The quantile-quantile plot for the overall fixed effects meta-analysis shows a deviation of the p values from in the tail of the plot (Figure 5-2). The fixed effects meta-analysis detected signals in the *ADAMTS7/MORF4L1/CHRNB4* at genome wide significance (Figure 5-2, Figure 5-3 and Table 5-7). There are two distinct signals in this region marked by rs7173743 and rs4380028 (Table 5-8 and Figure 5-5). The top SNP rs11072811 is in linkage disequilibrium with rs7173743 that has been reported at genome wide significance for CAD[43]. A forest plot of individual study estimates for rs11072811 is given in Figure 5-4. Rs11634042 in linkage with rs4380028, which has also been reported at genome wide significance for CAD, was detected at 5.73E-08 (Table 5-7 and Figure 5-5). Other SNPs which did not achieve genome wide significance but were detected at a threshold less than 1E-05 were found in or close to *LY75*, the 9p21 region*, CHRNB4*, *MSRA, CXCL12*, *FGFR*3, *PON2, ACVR2A, TNFRSF21, CD2AP, SPRY1, KCNE4, ACSL3, MTCH2, CELF1, STIM1* and *APOB* (Table 5-7)*.*

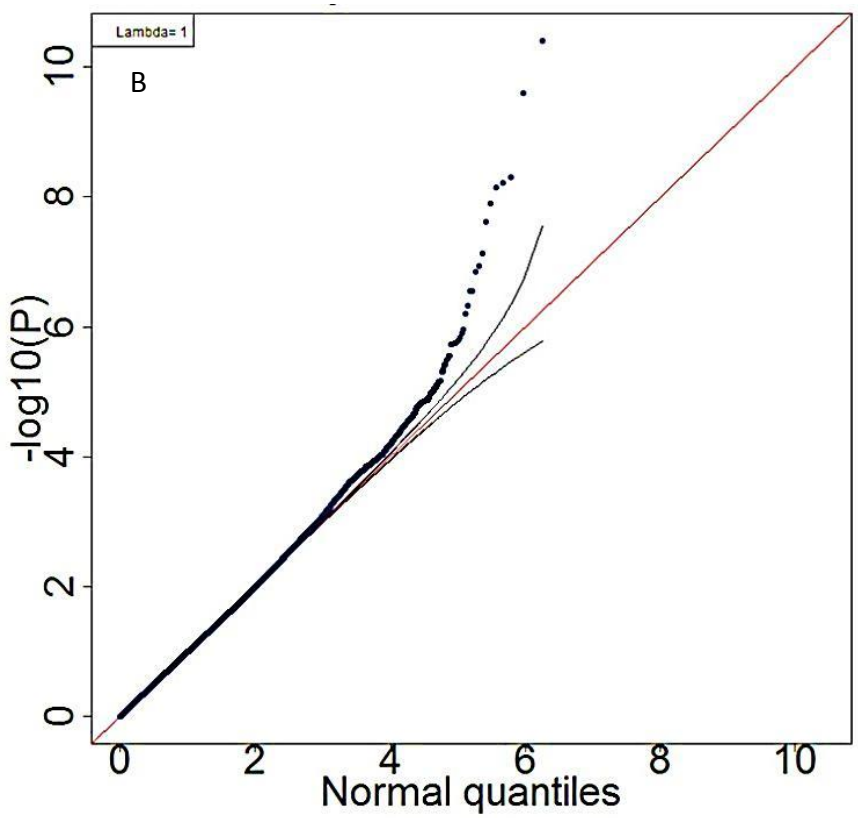

Figure 5-2: A – Quantile-quantile plot of p values from a meta-analysis of 14 genome wide association studies for coronary artery disease in type 2 diabetic populations. **B**- QQ plot of p values from a meta-analysis of 14 genome wide association studies and 8 CardioMetabochip studies for coronary artery disease in type 2 diabetic populations

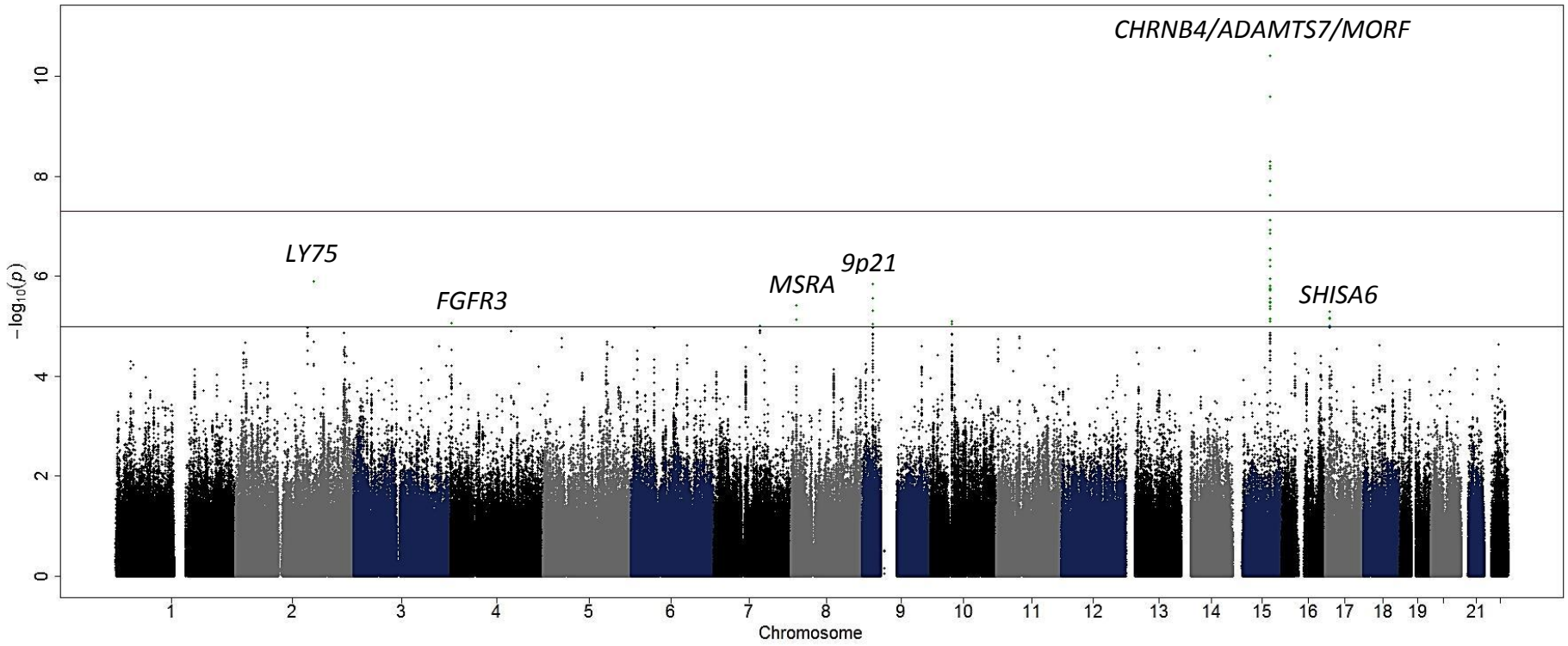

Figure 5-3: A Manhattan plot of from a meta-analysis of 14 genome wide association studies and 8 CardioMetabochip studies for coronary artery disease in type 2 diabetic populations. Green dots indicate suggestive signals.

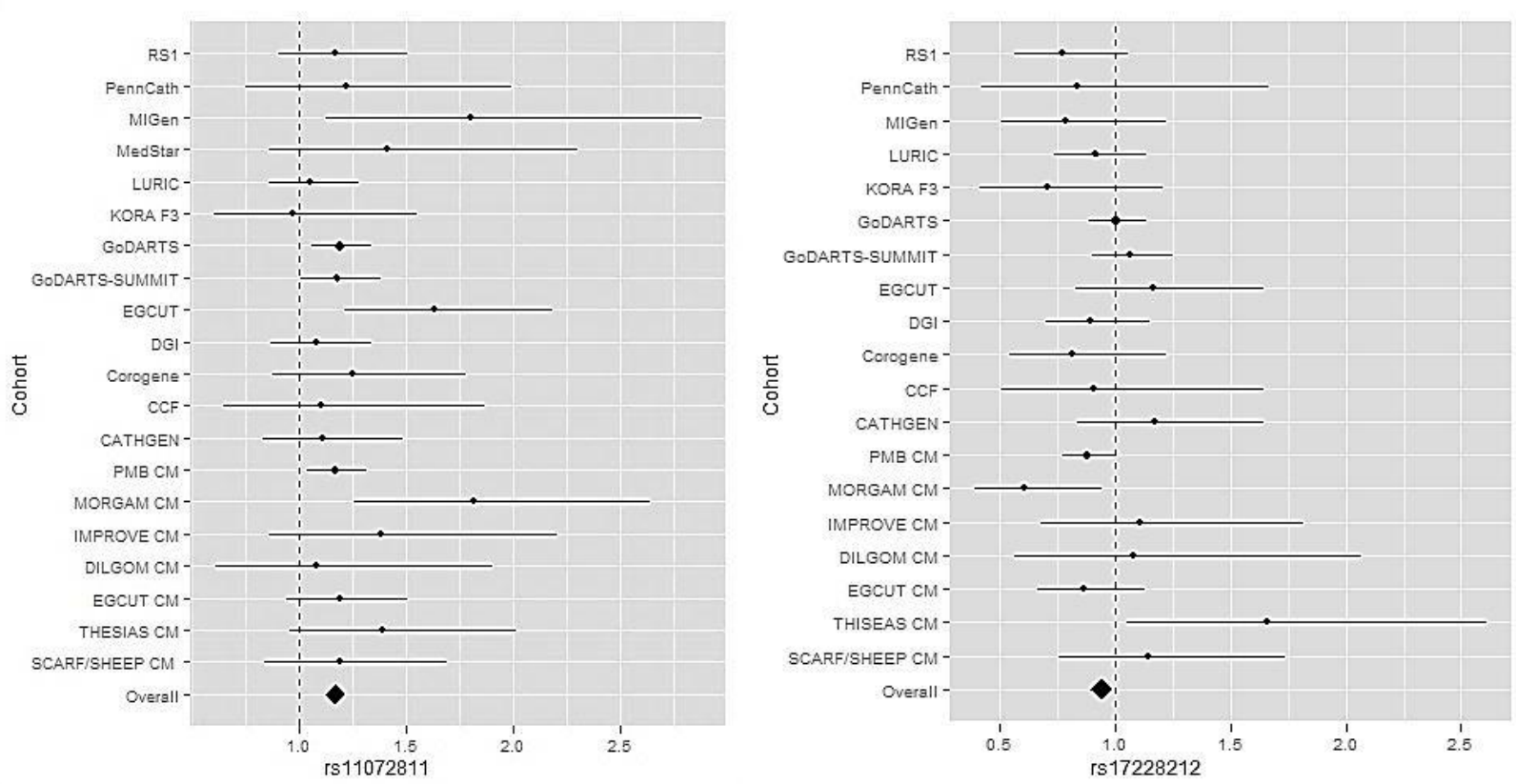

Figure 5-4: Forest plots of the top hit for *ADAMTS7* rs11072811 that reached genome significance for risk of coronary artery disease in type 2 diabetic individuals and of rs17228212 in *SMAD3* which has been previously associated with CAD but does not replicate in T2D individuals despite power to detect an association.

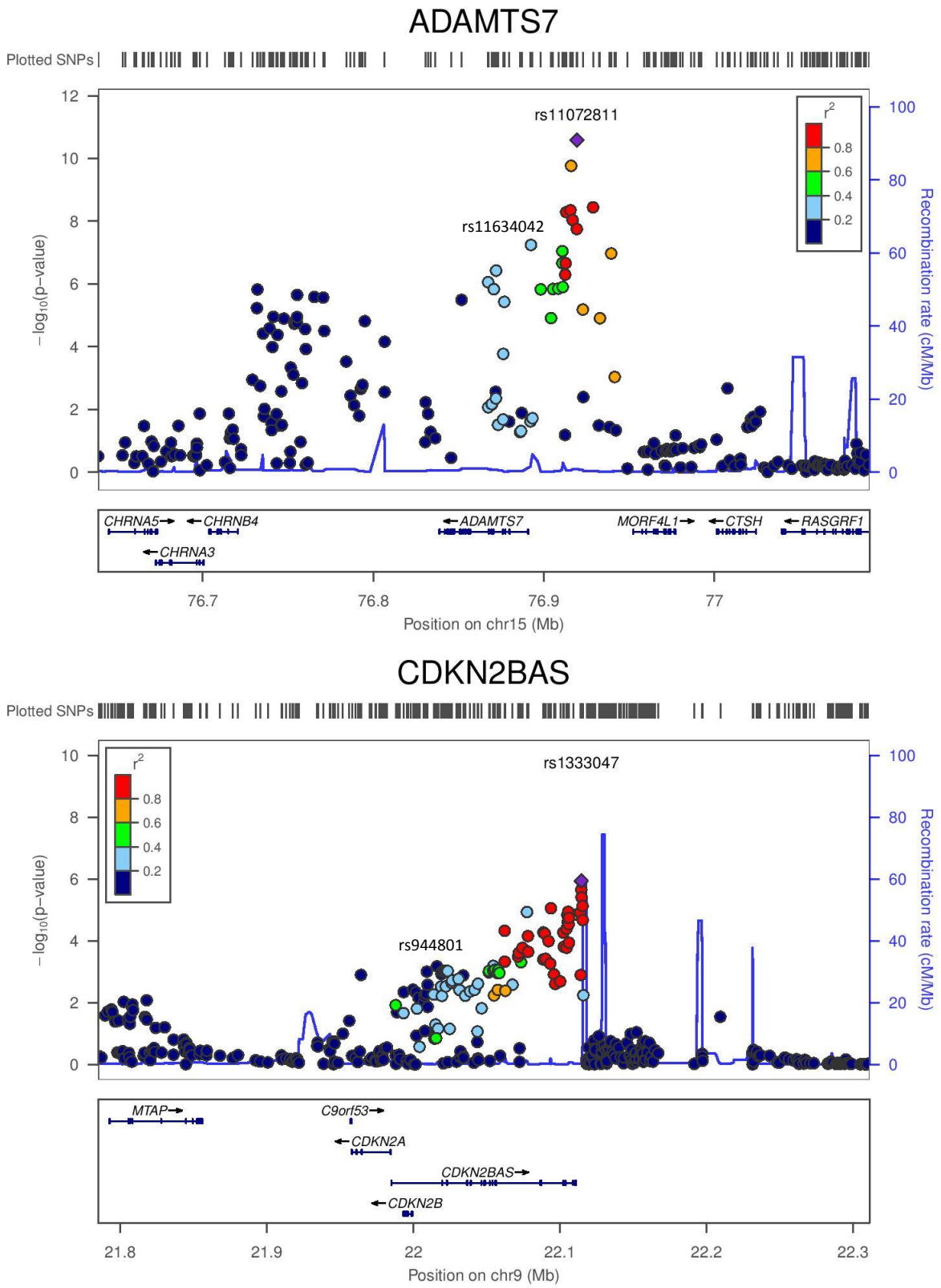

Figure 5-5: Locus zoom plots of *ADAMTS7* and *CDKNBAS* base on the p values obtained in a meta-analysis of coronary artery disease in type 2 diabetic individuals

Table 5-7: Top hits from the combined meta-analysis of CAD in T2D pruned for LD at p<1E-05

CHR	BP	SNP	EAF**	EA*	OR†	SE¥	P	i2	Gene
15	79132330	rs11072811	0.47	A	0.83	0.022	3.89E-11	0.00	*MORF4L1/ ADAMTS7*
15	76892405	rs11634042	0.42	T	0.85	0.024	5.73E-08	0.18	*MORF4L1/ ADAMTS7*
2	160729820	rs10497206	0.05	G	0.64	0.054	1.26E-06	0.56	*LY75-CD302/ PLA2R1*
9	22124504	rs1333047	0.53	A	0.88	0.023	1.44E-06	0.38	*CDKN2B-AS1/ UBA52P6/ DMRTA1*
15	78945040	rs8023822	0.23	C	1.17	0.037	1.86E-06	0.00	*CHRNB4/ CHRNA3/ CHRNA5*
8	10123974	rs13250969	0.44	C	1.18	0.040	3.86E-06	0.00	*MSRA/ LINC00599/ MIR124-1/ PRSS51/PRSS55*
17	10917434	rs5002484	0.35	T	1.18	0.040	4.97E-06	0.35	*RPL15P21/ SHISA6*
10	44691241	rs1623851	0.20	G	0.85	0.029	8.04E-06	0.00	*LINC00619/ CXCL12*
4	1762287	rs732754	0.19	A	1.23	0.055	8.49E-06	0.00	*TACC3/FGFR3*
7	95072081	rs43043	0.17	T	1.22	0.052	9.77E-06	0.00	*PON2/PON3/ ASB4*
2	148472953	rs12053348	0.11	T	0.78	0.041	1.04E-05	0.00	*RNA5SP106/ ACVR2A*
6	47362214	rs6458567	0.39	T	0.84	0.032	1.04E-05	0.00	*TNFRSF21/ CD2AP*
4	125266797	rs672858	0.49	G	0.89	0.024	1.25E-05	0.00	*SPRY1/*

CHR	BP	SNP	EAF**	EA*	OR†	SE¥	P	i2	Gene
									TECRP2
2	223936738	rs1440072	0.05	C	1.39	0.098	1.38E-05	0.00	*KCNE4/ACSL3/ ATG12P2/ HIGD1AP4*
11	47699398	rs7930612	0.15	A	0.85	0.031	1.58E-05	0.42	*AGBL2/MTCH2/ FNBP4*
11	47496827	rs7102372	0.14	T	0.84	0.032	1.75E-05	0.33	*CELF1/RAPSN/ NDUFS3/ PTPMT1/ KBTBD4/ RNU5E-10P*
5	38412674	rs16903965	0.11	T	1.26	0.065	1.76E-05	0.00	*EGFLAM*
11	4111900	rs3794050	0.12	A	0.79	0.041	1.83E-05	0.00	*STIM1/ RPS29P20/ RRM1*
5	130463135	rs6874948	0.24	C	1.18	0.043	2.07E-05	0.00	*ARL2BPP4/ HINT1/LYRM7*
2	21112689	rs12710745	0.41	G	0.88	0.025	2.14E-05	0.00	*C2orf43/APOB*
22	30865337	rs9620997	0.20	A	1.19	0.047	2.34E-05	0.00	*SEC14L3/ MTFP1/ SDC4P*

*Effect allele; **Effect allele frequency; †Odds ratio; ¥Standard error

Table 5-8: Linkage disequilibrium relationships between known hits rs7173743 and rs4380028 and SNPs that reached genome wide significance in the coronary artery disease in type 2 diabetic populations meta-analysis estimated from HapMap2.

SNPs		**rs7173743** R^2/D'	**rs4380028** R^2/D'	**rs11072811** R^2/D'	**rs7168915** R^2/D'	**rs4539564** R^2/D'	**rs8032771** R^2/D'	**rs8035039** R^2/D'	**rs11072810** R^2/D'
rs7173743	Known		0.62/0.91	0.78/0.96	0.93/0.97	0.76/0.96	0.76/0.96	0.73/0.96	0.78/0.96
rs4380028	Known			0.51/0.90	0.67/0.96	0.55,0.95	0.55/0.95	0.55/0.95	0.51/0.90
rs11072811	Meta				0.78/0.96	0.97/1.00	0.97/1.00	0.97/1.00	1.00/1.00
rs7168915	Meta					0.82/1.00	0.97/1.00	0.82/1.00	0.78/0.96
rs4539564	Meta						1.00/1.00	1.00/1.00	0.97/1.00
rs8032771	Meta							1.00/1.00	0.97/1.00
rs8035039	Meta								0.97/1.00
rs11072810	Meta								

5.3.6 Random effects model

The random effects model was applied to 17904 SNPs where significant heterogeneity was detected in the effect estimates. After removing extreme outliers from MedStar and PennCath there was no deviation in the mean lambda estimate and the lambda for heterogeneity of 10.19 was used to correct the random effects model. The final analysis did not reveal any signals at genome wide significance (Table 5-9); however the top hit was rs2891168, a proxy for rs4977574 and rs1333049 in the 9p21 region that has been reported at genome wide significance for CAD (Figure 5-6). The same risk allele was reported for rs2891168 that has been reported for rs4977574 and rs1333049. Based on estimates from this study we found the random effects model (RE2) proposed by Han and Eskin to be more powered to detect an association than the traditional RE model (Table 5-9).

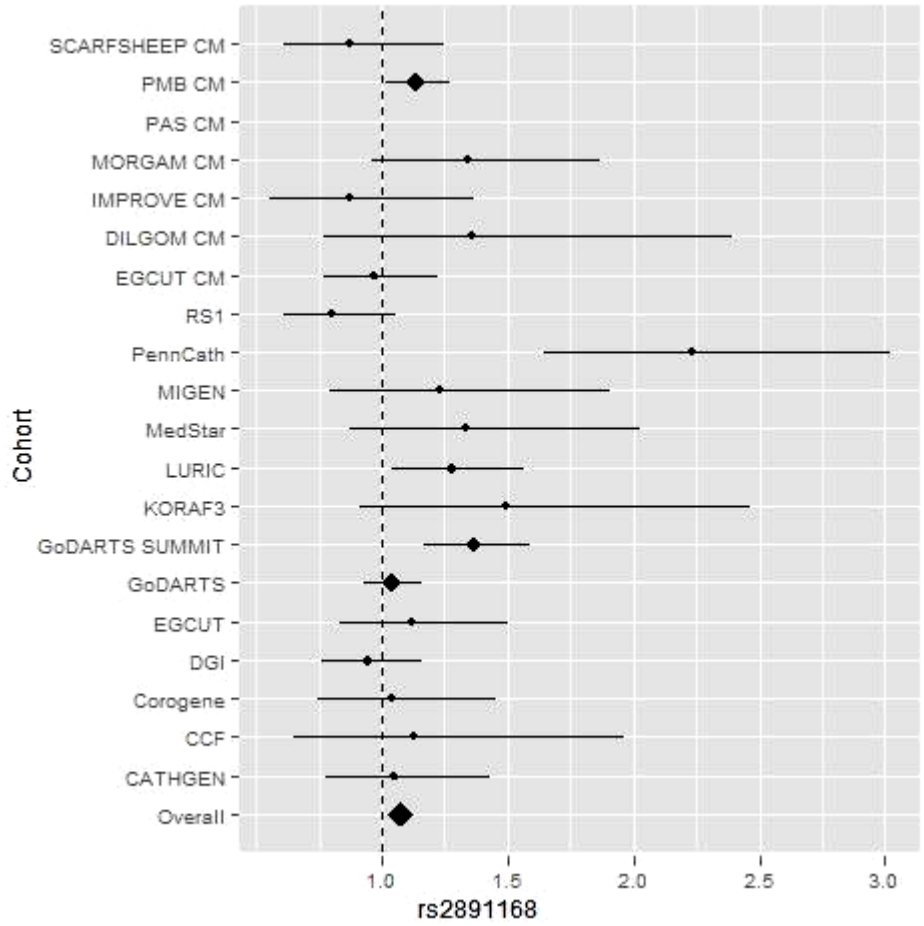

Figure 5-6: A Forest plot of the top hit from the Han and Eskin random effects model. Rs2891168 in the 9p21 region has heterogeneous allelic effects amongst the cohorts included in the meta-analysis

Table 5-9: Random effects models all SNPs with a p value of less than 1E-03 for the random effects model proposed by Han and Eskin from the combined analysis

Gene	SNP	Studies	EA*	EAF**	$P_{FE†}$	OR_{FE}	SE_{FE}	$P_{RE¥}$	OR_{RE}	SE_{RE}	$P_{RE2Φ}$	I^2	P_Q
CDK2NBAS	rs2891168	19	G	0.47	4.89E-04	1.09	0.0267	3.36E-02	1.08	0.0402	8.05E-04	41.38	3.11E-02
ASTN2	rs2210769	10	T	0.12	6.75E-04	1.23	0.0615	1.05E-01	1.21	0.1177	8.34E-04	66.90	1.30E-03
ARL2BPP6	rs7716149	10	G	0.82	6.64E-04	1.13	0.0404	5.81E-03	1.17	0.0669	9.47E-04	53.28	2.31E-02
ARL2BPP6	rs11748167	10	G	0.80	6.87E-04	1.13	0.0404	5.90E-03	1.17	0.0669	9.79E-04	53.24	2.32E-02

*Effect allele; **Effect allele frequency; $_{†}$FE ~ fixed effects p value (P), odds ratio (OR) and standard error (SE); $_{¥}$RE ~ random effects p value (P), odds ratio (OR) and standard error (SE); $_{Φ}$Han and Eskin Random effects model p value

5.3.7 Comparison of effects estimated form the meta-analysis with known coronary artery disease loci and Type 2 diabetes loci

5.3.7.1 Coronary artery disease

We detected an association at p less than 0.05 for 23 of the CAD loci (Table 5-10). The risk allele from the current meta-analysis was consistent with the published risk allele for 22 of the 23 significant signals (Figure **5-7**A). The power calculations identified three loci that we had the power to detect for the published effects: *SMAD3* (rs17228212)*, VEGFA* (rs6905288) and *ABO* (rs514659). Rs17228212 in *SMAD3* had an inverse effect on CAD risk when compared to the published estimate at a p value of 3.99E-02 (Table 5-10 and Figure **5-7**). We detected no association with *ABO* and *VEGFA* in this study.

The 23 significant loci, *VEGFA* and *ABO* were tested for heterogeneity in allelic effects observed in this meta-analysis compared to the estimates obtained in chapter 4. Rs7173743 in *ADAMTS7* had heterogeneous allelic effects when compared to published estimates (p=5.60E-03) and had a greater effect on CAD in T2D (Table 5-10). The well-known CAD signal rs1333049 in *CDK2NBAS* was also significantly heterogeneous (p=4.15E-04) and had a smaller allelic effect on CAD in T2D. Rs17228212, in *SMAD3*, also had highly heterogeneous allelic effects (p=1.00E-06) where the published risk allele has a nominally protective effect on CAD in T2D (Table 5-10). Rs6905288 in *VEGFA* also displayed heterogeneous allelic effects and was found not to be associated with CAD in T2D by this study and while we found no heterogeneity in the effect estimates for *ABO* we also do not detect an association. Rs7865618, in *CDKN2BA* and rs4380028 in *ADAMTS17*, also showed heterogeneity in effect estimates that was nominally significant.

Table 5-10: Association of previously reported coronary artery disease loci with coronary artery disease in this meta-analysis

SNP	Gene	EA*	EAF**	OR† published	OR study	P	N studies	N samples	$P_{heterogeneity}$
rs7173743	*ADAMTS7*	T	0.57	1.07	1.15	3.62E-09	19	16564	5.60E-03
rs4380028	*ADAMTS7*	C	0.62	1.07	1.13	1.49E-06	19	16593	5.00E-02
rs1333049	*9p21*	C	0.47	1.23	1.11	2.14E-05	21	16968	4.15E-04
rs1746048	*CXCL12*	C	0.86	1.09	1.15	4.27E-05	18	15803	
rs501120	*CXCL12*	A	0.86	1.07	1.14	1.84E-04	18	15541	
rs515135	*APOB*	G	0.82	1.08	1.12	2.65E-04	17	15480	
rs1122608	*LDLR*	G	0.75	1.10	1.10	4.81E-04	20	16907	
rs1561198	*VAMP5-VAMP8- GGCX*	A	0.46	1.05	1.10	7.17E-04	20	16912	
rs7865618	*9p21*	A	0.57	1.18	1.09	9.25E-04	20	16910	1.20E-02
rs1412444	*LIPA*	T	0.33	1.09	1.12	1.85E-03	11	9785	
rs602633	*SORT1*	C	0.78	1.12	1.10	2.04E-03	19	16413	
rs6725887	*WDR12*	C	0.12	1.12	1.13	2.95E-03	18	15801	
rs646776	*CELSR2*	T	0.77	1.14	1.09	3.15E-03	19	16425	
rs2023938	*HDAC9*	G	0.10	1.07	1.14	3.16E-03	16	15282	
rs9982601	*KCNE2*	T	0.13	1.13	1.12	4.12E-03	17	15479	
rs9369640	*PHACTR1*	A	0.63	1.09	1.07	8.53E-03	20	16906	
rs12526453	*PHACTR1*	A	0.67	1.09	1.06	2.08E-02	20	16912	

SNP	Gene	EA*	EAF**	OR† published	OR study	P	N studies	N samples	$P_{heterogeneity}$
rs17465637	*MIA3*	C	0.72	1.14	1.07	2.99E-02	17	14277	
rs11556924	*Z3HC1*	C	0.62	1.09	1.06	3.60E-02	21	16976	
rs12190287	*TCF21*	C	0.60	1.07	1.06	3.78E-02	20	16907	
rs4845625	*IL6R*	T	0.44	1.04	1.06	3.83E-02	20	16655	
rs17228212	*SMAD3*	C	0.27	1.21	0.94	3.99E-02	21	16979	1.00E-06
rs2895811	*HHIPL1*	C	0.43	1.06	1.06	4.88E-02	19	16592	
rs2246833	*LIPA*	T	0.34	1.06	1.05	7.08E-02	20	16913	
rs2505083	*KIAA1462*	C	0.43	1.06	1.05	8.67E-02	20	16909	
rs264	*LPL*	T	0.14	1.05	0.93	9.79E-02	17	15207	
rs7651039	*BTG*	C	0.51	1.06	0.96	1.15E-01	19	16535	
rs4252120	*PLG*	T	0.72	1.06	1.05	1.21E-01	20	16893	
rs17514846	*FURIN-FES*	A	0.45	1.05	1.04	1.26E-01	17	15137	
rs579459	*ABO*	C	0.21	1.10	1.05	1.35E-01	19	16593	
rs17464857	*MIA3*	T	0.86	1.05	1.06	1.42E-01	17	15605	
rs964184	*ZNF259, APOA5, APOA4, APOC3, APOA1*	G	0.13	1.13	1.07	1.54E-01	13	10859	
rs10933436	*INPP5*	A	0.44	1.06	1.04	1.56E-01	17	15516	
rs12205331	*ANKS1A*	C	0.77	1.04	1.05	1.64E-01	15	14851	
rs17609940	*ANKS1A*	G	0.77	1.07	1.05	1.70E-01	13	10863	
rs2954029	*TRIB1*	A	0.53	1.04	1.03	1.95E-01	20	16656	

SNP	Gene	EA*	EAF**	OR† published	OR study	P	N studies	N samples	$P_{heterogeneity}$
rs3184504	*SH2B3*	T	0.48	1.07	1.04	1.96E-01	17	15659	
rs2075650	*APOE-APOC1*	G	0.15	1.11	1.06	2.00E-01	13	13095	
rs6905288	*VEGFA*	T	0.56	1.23	1.03	2.14E-01	19	16570	1.00E-03
rs4773144	*COL4A1, COL4A2*	G	0.42	1.07	1.03	2.38E-01	21	16955	
rs688034	*SEZ6L*	T	0.34	1.11	0.96	2.82E-01	14	10929	
rs974819	*PDGFD*	A	0.27	1.07	1.03	2.82E-01	17	15962	
rs2252641	*ZEB2-AC074093.1*	G	0.43	1.04	1.03	3.32E-01	20	16915	
rs9319428	*FLT1*	G	0.32	1.05	1.03	3.35E-01	19	16523	
rs9818870	*MRAS*	T	0.15	1.07	1.03	3.82E-01	18	15802	
rs4937126	*ST3GAL4*	G	0.70	1.06	0.97	3.87E-01	11	9781	
rs7692387	*GUCY1A3*	G	0.80	1.06	1.03	4.01E-01	17	13548	
rs11206510	*PCSK9*	T	0.83	1.08	1.03	4.18E-01	20	16098	
rs273909	*SLC22A4-SLC22A5*	C	0.13	1.09	1.03	4.22E-01	18	15469	
rs15563	*UBE2Z, GIP, ATP5G1, SNF8*	C	0.52	1.04	1.02	4.66E-01	18	15688	
rs12936587	*RASD1, SMCR3, PEMT*	G	0.56	1.04	1.02	4.73E-01	17	15963	
rs6544713	*ABCG5-ABCG8*	T	0.31	1.06	1.02	4.86E-01	15	14886	
rs514659	*ABO*	C	0.36	1.21	1.02	5.22E-01	19	16592	
rs204832	*LPA*	G	0.72	1.06	1.03	5.33E-01	12	10545	
rs1878406	*EDNRA*	T	0.13	1.06	1.03	5.37E-01	16	13314	

SNP	Gene	EA*	EAF**	OR† published	OR study	P	N studies	N samples	$P_{heterogeneity}$
rs3739998	*KIAA1462*	C	0.44	1.15	1.02	5.63E-01	13	10859	
rs10947789	*KCNK5*	T	0.75	1.06	0.98	5.72E-01	19	16593	
rs365302	*FNDC1*	C	0.23	1.11	1.02	6.28E-01	13	10868	
rs12413409	*CYP17A1, CNNM2, NT5C2*	G	0.91	1.12	1.02	6.32E-01	17	15465	
rs9326246	*ZNF259, APOA5, APOA4, APOC3, APOA1*	C	0.07	1.09	1.02	6.51E-01	15	14955	
rs6922269	*MTHFD1L*	A	0.26	1.23	1.01	6.68E-01	18	15984	
rs16893526	*FAM46A –IBTK*	G	0.92	1.13	0.98	6.70E-01	16	15146	
rs12539895	*7q22*	A	0.21	1.08	0.99	6.85E-01	18	16361	
rs17114036	*PPAP2B*	A	0.91	1.11	1.02	7.03E-01	15	14952	
rs2259816	*HNF1A*	T	0.36	1.08	1.01	7.10E-01	20	16908	
rs2943634	*KIAA1486*	C	0.68	1.21	1.01	7.68E-01	21	16971	
rs7808424	*ASZ1*	G	0.10	1.10	1.01	9.13E-01	12	10664	
rs1231206	*SMG6*	A	0.34	1.07	1.00	9.31E-01	12	10545	

*Effect allele; **Effect allele frequency; †Odds ratio

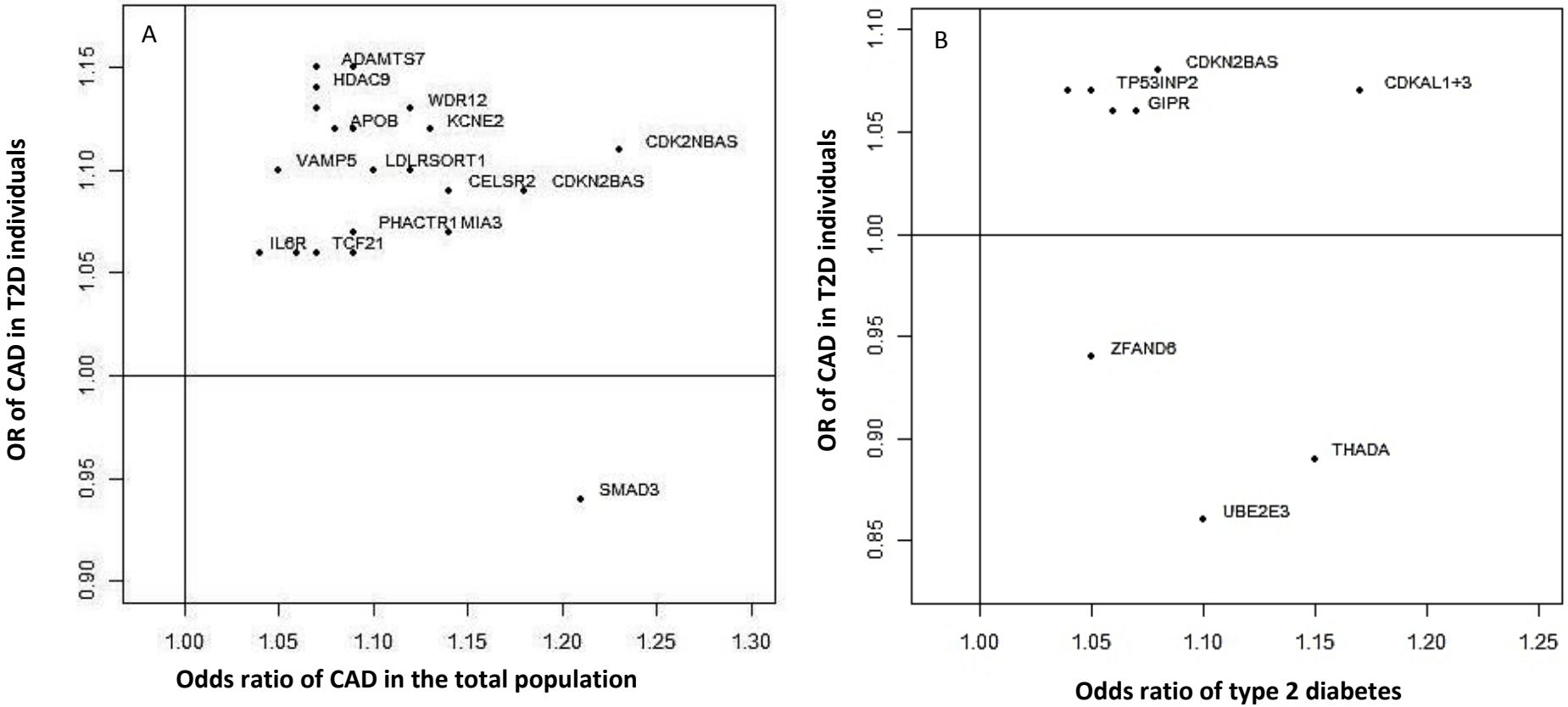

Figure 5-7: A - Comparison of odds ratios from coronary artery disease in type 2 diabetic individuals with published odds ratios estimated from mixed non-diabetic and diabetic populations. **B** - Comparison of the published odds ratios for known type 2 diabetes genes with OR for CAD if the SNP was associated with CAD at $p<0.05$

5.3.7.2 Type 2 diabetes

Of the 65 T2D loci tested 9 SNPs were also associated with CAD at a threshold of 0.05 (Table 5-11). Six of these SNPs in the 9p21 region, *TTP53INP2*, *TP53INP1, PEPD, CDKAL3* and *CDKAL1* shared the same risk allele with CAD and T2D (Table 5-11 and Figure **5-7**B). The odds ratio for CAD and T2D was identical for rs944801 in the 9p21 region but the SNP was not in linkage with other SNPs reported for CAD in the 9p21 region (Table 5-12 and Figure 5-5). However the previously established T2D SNP in the 9p21 region, rs10811661, was not associated with CAD in this study (OR=1.03 for the C allele, p=0.38). The published risk allele is the major T allele and the OR is 1.20 for T2D[298]. We also found that the risk allele for CAD and T2D were inverted for SNPs in *ZFAND6, THADA* and *UBE2E3* (Table 5-11 and **Figure 5-7**B).

Table 5-11: Comparison of allelic odds ratios obtained from the meta-analysis of coronary artery disease in type 2 diabetes individuals, presented here, compared to odds ratios from the DIAGRAM type 2 diabetes meta-analyses for diabetes associated loci

RSID	EA*	EAF**	CAD OR†	CAD P	T2D OR¥	T2D P	Gene
rs944801	C	0.56	1.08	3.87E-03	1.08	2.42E-09	*9p21*
rs896854	T	0.51	1.07	1.37E-02	1.05	2.13E-05	*TP53INP2*
rs7845219	T	0.52	1.06	1.71E-02	1.06	4.57E-06	*TP53INP1*
rs7612463	C	0.89	0.86	1.86E-02	1.10	9.79E-04	*UBE2E3*
rs8182584	T	0.38	1.07	2.29E-02	1.04	2.20E-03	*PEPD*
rs9368222	A	0.30	1.07	2.57E-02	1.17	7.04E-34	*CDKAL3*
rs7756992	G	0.30	1.07	2.83E-02	1.17	6.95E-35	*CDKAL1*
rs11634397	G	0.66	0.94	3.00E-02	1.05	1.35E-04	*ZFAND6*
rs11899863	C	0.92	0.89	3.97E-02	1.15	9.48E-11	*THADA*

*Effect allele; **Effect allele frequency; †Coronary artery disease odds ratio; ¥Type 2 diabetes odds ratio

5.3.8 Testing for independent effects in the 9p21 region

The lead SNP for CAD rs1333049[38] is not in linkage with the two T2D signals rs944801 and rs10811661[266] in the 9p21 region. Rs944801 and rs10811661 are also independent signals for T2D (Table 5-12). While rs944801 was reported in the meta-analysis of CAD in T2D

individuals the SNP was not available for a subset of GoDARTS individuals typed on the CardioMetabochip so rs7030641 (rs944801, R^2=0.966, D'=1) a proxy for rs944801 was used to represent this SNP.

First we replicated the known associations of rs1333049 with CAD and rs7030641 (rs944801) and rs10811661 with T2D in GoDARTS (Table 5-13). Rs1333049 minor allele (G) was associated with an increased risk of CAD (odds ratio=1.15) irrespective of T2D status as has been previously published. The major T allele of rs7030641 (odds ratio=1.11) and the major T allele of rs10811661 (OR=1.17) were associated with increased risk of T2D, consistent with the published associations.

We then investigated whether the SNPs interacted with T2D status to affect the risk of CAD. We found a nominally insignificant interaction for rs7030641 (rs944801) T allele with diabetic status, where the odds ratio of the interaction was 1.24 for diabetes status and the p value for interaction was 8.19E-02 in GoDARTS. Rs1333049 and rs10811661 did not interact with diabetes status and CAD in GoDARTS. The association of the SNPs was tested just in the T2D subgroup and we found that rs1333049 and rs7030641 (rs944801) were both associated with CAD in T2D. We observed that the allelic effect of rs7030641 (rs944801) on CAD was greater in the T2D population of GoDARTS compared to the allelic effect estimated from a GoDARTS population that included diabetics and non-diabetics – 1.17 vs. 1.11. This difference in effects is supported by the nominally insignificant result we observed for an interaction for this SNP with T2D status in GoDARTS. We were unable to compare the allelic effects observed for rs7030641 with CAD in GoDARTS to the allelic effects observed in the CARDIoGRAM+C4D meta-analysis as they were not available. Consistent with published findings rs10811661 was not associated with CAD in the general or T2D populations from GoDARTS (Table 5-13).

When all three SNPs were included in the logistic regression model none of them were associated with CAD (p>0.05) and the OR of rs1333049 was reduced from 1.16 for CAD in T2D to 1.08 in the three SNP model (Table 5-13) in T2D individuals. These results indicate that rs1333049 is associated with CAD irrespective of T2D status and confirms previous findings that rs10811661 is not associated with CAD irrespective of T2D status. Rs944801

has been identified as a novel locus for CAD in T2D and interacts with T2D status to confer a higher risk of CAD in T2D individuals (Figure 5-8).

Table 5-12: The 9p21 region is a locus for coronary artery disease (CAD) and type 2 diabetes (T2D). This table shows the linkage disequilibrium relationships amongst the reported signals for CAD and T2D.

Disease	**SNP**	**rs7044859** R^2/D'	**rs1333049** R^2/D'	**rs3217992** R^2/D'	**rs944801** R^2/D'	**rs10811661** R^2/D'
Coronary artery disease	rs7044859		0.145/0.395	0.593/0.905	0.577/0.985	0.001/0.038
Coronary artery disease	rs1333049			0.301/0.667	0.256/0.639	0.009/0.179
Coronary artery disease	rs3217992				0.395/0.966	0.004/0.133
Type 2 diabetes	rs944801					0.01/0.142
Type 2 diabetes	rs10811661					

Table 5-13: The association of coronary artery disease signal rs1333049 and type 2 diabetes signals rs10811661 and rs944801 in the 9p21 region with CAD in T2D individuals in the GoDARTS population.

SNP	Model	EA*	EAF**	Diabetes			Coronary artery disease			Interaction with diabetes		
				OR[†]	95%CI[¥]	P	OR	95%CI	P	OR	95%CI	P
rs1333049	All	G	0.48	1.05	(0.97-1.13)	1.78E-01	1.15	(1.02-1.29)	2.11E-02	1.07	(0.83-1.38)	5.44E-01
rs944801[Φ]	All	T	0.57	1.09	(1.01-1.18)	1.49E-02	1.11	(0.98-1.24)	8.82E-02	1.24	(0.96-1.60)	8.19E-02
rs10811661	All	T	0.85	1.17	(1.06-1.29)	9.32E-04	0.90	(0.77-1.05)	1.81E-01	0.84	(0.54-1.09)	3.29E-01
rs1333049	T2D	G	0.47				1.16	(1.01-1.33)	4.22E-02			
rs944801[Φ]	T2D	T	0.59				1.17	(1.02-1.34)	4.07E-02			
rs10811661	T2D	T	0.85				0.86	(0.70-1.04)	1.30E-01			
rs1333049	T2D 3 SNPs	G	0.47				1.08	(0.90-1.29)	3.52E-01			
rs944801[Φ]	T2D 3 SNPs	T	0.59				1.14	(0.95-1.36)	1.32E-01			
rs10811661	T2D 3 SNPs	T	0.85				0.85	(0.70-1.04)	1.18E-01			

*Effect allele; **Effect allele frequency; [†]Odds ratio; [¥]95% confidence intervals; [Φ]Represented by rs7030641 (R^2=0.966, D'=1) in GoDARTS

Figure 5-8: Rs1333049 is an established SNP for coronary artery disease (CAD) and is not associated with type 2 diabetes (T2D) and has a significantly lower effect in T2D populations. Rs10811661 is an established SNP for type 2 diabetes and is not associated with CAD however a signal marked by rs944801 is associated with both CAD and T2D and has a stronger effect on CAD in T2D individuals

5.4 Discussion

Here we report the largest genetic study to date to identify common variants that contribute to CAD in T2D individuals. We combined genetic data from 6995 CAD cases and 11203 controls and identified two independent signals in *ADAMTS7* that are associated with CAD in T2D at genome wide significance. Other suggestive signals were detected in *LY75*, the 9p21 region, *CHRNB4*, *MSRA, CXCL12*, *FGFR*3, *PON2/3, ACVR2A, TNFRSF21, CD2AP, SPRY1, KCNE4, ACSL3, MTCH2, CELF1, STIM1* and *APOB*. A comparison of within study allelic effects with published effects for known CAD loci revealed unexpected deviations. The allelic effects of variants in *ADAMTS7* were inflated compared to published estimates while allelic effects in the 9p21 region were deflated and the allelic effects for *SMAD3* were inverted. Despite the power to

detect an association with *VEGFA* and *ABO* these loci were not associated with CAD in T2D in this study. No overlap between diabetes loci and CAD risk loci has been reported by large genetic studies but in this study we found an overlap in signal in the 9p21 region.

5.4.1 Top hits and novel signals detected by the fixed effects model

ADAMTS7 is a member of the metalloproteinase family of a disintegrin and metalloproteinase with thrombospondin motifs. Studies in rats have shown that overexpression of this gene increases migration and invasion of the endothelium by vascular SMCs and increases the thickness of injured arteries[299]. Two independent signals represented by rs7173743 and rs4380028 had been published for *ADAMTS7* for CAD in diabetic and we detect proxies for these SNPs at genome wide significance in this study[43, 48].

We detected suggestive novel signals in *CXCL12, SPRY1* and *FGFR3* that have also been implicated in plaque formation and structure. Signals in *CXCL12* have been reported from large genetic studies of CAD[38, 48]. Studies in rabbits have revealed that LDL-C induces the expression of *CXCL12* in endothelial cells that increases monocyte adhesion to atherosclerotic plaque extremities[61]. The expression of *FGFR*3 expression is restricted to within atherosclerotic plaques[300] where it acts as a receptor for *FGF-1* and *FGF-2,* pleiotropic growth factors that affect cell types involved in restenosis and atherosclerosis[301]. *SPRY1* is involved in the negative regulation of angiogenesis that occurs in the final stages of atherosclerosis[302]. Other suggestive signals affect atherosclerotic risk factors such as central obesity and oxidative stress.

Two separate signals in *MSRA* have been associated with central obesity and with rheumatoid arthritis[247, 303], while the signal detected in this study is not linked to either of the reported signals, the gene has been implicated for central obesity a risk factor for CAD and T2D[247]. *MTCH2* is another gene that is associated with BMI[304, 305] which is also in the top set of genes detected in this meta-analysis. Increased serum levels of *PON2/3* have been positively correlated with pro-inflammatory markers in CAD patients[306]. Within the liver and vasculature, paraoxonases such as *PON3* protect the body from atherosclerosis through their antioxidant activity on HDL-C and from premature apoptosis[307-310]. *CHRNB4/CHRNA3/CHRNA5* have all been associated with nicotine dependence and increased smoking[194, 195], a well-established risk

factor for cardiovascular disease. *APOB* and *ACSL3* are involved with lipid metabolism[311, 312]. Other signals were observed in genes that control cellular processes and have a direct effect on the heart.

ACVR2A, TNFRSF21, CD2AP are involved in cell proliferation, differentiation and apoptosis. *CD2AP* is specifically involved in actin remodelling and membrane trafficking, and contains a signal for Alzheimer's disease[313]. *KCNE4* has a role in the regulation of heart rate[314] while CELF1 and STIM1 have been associated with Myotonic dystrophy[315] and cardiac hypertrophy[316]. These signals may have an effect on CAD through their action on cells in the atherosclerotic plaques or directly on the heart muscle itself increasing the risk of a CAD event. A new random effects model which has been proposed by Han and Eskin has been tested for the first time on summary statistics from a large population.

5.4.2 Evaluation of the Han and Eskin random effects model

Based on the findings from this study the random effects model proposed by Han and Eskin, 2012 has increased power to detect significant signals when compared to the traditional random effects model. The new random effects model does not assume heterogeneity under the null hypothesis and partitions heterogeneity into heterogeneity of allelic effects from heterogeneity due to other factors like population stratification. These features and the performance of the new random effects model in this meta-analysis would indicate that it is more powered to detect an effect than traditional random effects models.

5.4.3 Known CAD signals in the context of CAD in T2D populations

We detected 23 of the known CAD SNPs at nominal significance ($p<0.05$) with allelic effects consistent with published allelic estimates for 17 SNPs and heterogeneous allelic effects for 5 loci. Power calculations also identified SNPs in *VEGFA* and *ABO* which we had the power to detect but we did not observe a significant association with CAD in this analysis. Significant ($p<0.05$) heterogeneity in allelic effects was detected between published estimates and effects estimated in this study for the genome wide significant signals in *ADAMTS7,* for two signals in the 9p21 region and for a signal in *SMAD3.*

The CAD signal in *SMAD3* has the opposite risk allele compared to published data and the allelic effect is nominally significant ($p<0.05$)[48]. *SMAD3* is directly involved in atherosclerotic processes as it is directly phosphorylated by *TGF-β,* which allows it to bind *SMAD4* and travel to the nucleus where the complex regulates the expression of target genes[317]. In atherosclerotic regions *TGF-β* acts specifically on smooth muscle cells in fibrous plaques and macrophages in streaks/fatty lesions[318] but *SMAD* signalling can also be activated through the *RAGE* and *MAPK/38* pathways to produce vascular sclerosis[319]. The inverted allelic effects for the reported SNP could be due to a number of different factors.

Thiazolidinedione (TZD) drugs are routinely prescribed to improve insulin sensitivity in T2D individuals. TZDs exert their effects by activating peroxisome proliferator-activated receptors, in particular *PPAR y,* to alter the expression of genes related to glucose and lipid metabolism[320]. While some TZDs like rosiglitazone and repaglinide have been removed from the market because trials have shown an increased risk of cardiovascular disease, pioglitazone has been shown to reduce the risk of all cause death, myocardial infarction, stroke[321] and reduces the size of atherosclerotic plaques[322]. Pioglitazone inhibits the expression of *TGF-β* in the kidneys[323, 324] and protects diabetic individuals against nephropathy[323, 324]. It may be that the suppression of *TGF-β* negates the effect of the risk allele of *SMAD3* in the T2D population but further investigations will have to be carried out to verify this hypothesis.

The effect of the risk alleles for both signals in *ADAMTS7* on CAD is much larger in T2D individuals but is still directionally consistent with the published estimates but the reason for this is not apparent. We also detected heterogeneous allelic effects for *VEGFA* and in the 9p21 region. The lack of association of SNPs *VEGFA* with CAD in T2D may be due to an interaction with the general obesity observed in T2D populations as this locus has also been associated with waist-hip ratio. However, the heterogeneity observed in the 9p21 region is not easily explained. Results from this study have indicated that there is complex genetic architecture across the 9p21 region with respect to its association with CAD and T2D.

5.4.4 Rs944801 is determinant of both T2D and CAD

The literature to date indicates that signals in the 9p21 region, rs10811661 for T2D and rs1333049 for CAD, do not overlap[325]. We confirmed that rs1333049 was associated with CAD in this study but was not associated with diabetes and did not interact with diabetes status. We also found that rs10811661 was not associated with CAD and did not interact with T2D status. We did find an overlap in both T2D and CAD phenotypes for the new T2D signal rs944801 in the 9p21 region which has been associated with CAD and T2D, and increases the risk of CAD in T2D individuals in GoDARTS. These findings indicate that rs1333049 is purely a CAD signal; that rs10811661 is purely a T2D locus and rs944801 is both a CAD and T2D locus.

5.4.5 Concluding remarks

One of the weaknesses of the current study is the inability to detect loci that have heterogeneous allelic effects between diabetics and non-diabetics. The published analyses include populations that are not independent of the populations included in this meta- analysis therefore it is not possible to test for SNPs for an interaction with T2D status. Therefore we have requested identical non-diabetic analyses from participating cohorts so that we may formally test for loci that interact with T2D status. Another weakness of the study is that we only have estimates for the full population for SNPs that overlap between the GWAS data and the CardioMetabochip so we are not powered to detect SNPs from the GWAS data with small effect sizes. We plan to add additional data from cohorts that have GWAS data to refine some of the suggestive signals.

We have conducted the first and largest meta-analysis of CAD in T2D populations. We have detected associations of two previously published signals in *ADAMTS7* at genome wide significance with CAD in T2D. Several suggestive signals have also been identified in biologically relevant genes such as *PON2/3, MSRA, FGFR3 and LY75* that warrant further investigation. We have shown that rs944801 in the 9p21 region is associated with CAD in T2D and overlaps the T2D and CAD phenotypes. We have shown that *VEGFA* and *ABO* are not associated with CAD in T2D despite the power to detect a signal and that *SMAD3* has a protective effect against CAD in T2D. Loci with heterogeneous allelic effects will be identified through a formal diabetes

stratified analysis. As part of the SUMMIT programme these data will also be combined with proteomic and metabolomics biomarker data that is being generated for some overlapping samples. These data will form part of a combined analysis to identify new biomarkers for CAD in T2D populations.

Chapter 6: Single nucleotide polymorphisms directly influencing 25-hydroxyvitamin D levels have the expected effect on coronary artery disease events – a Mendelian randomization study in Go-DARTS

6.1 Introduction

Low vitamin D levels are common[326-329] and an extensive body of observational data supports an association between low 25-hydroxy vitamin D (25OHD) levels and higher risk of cardiovascular events including patients referred for coronary angiography, patients with type 2 diabetes and in the general population[326, 330, 331]. Establishing a causal relationship between vitamin D status and cardiovascular disease is difficult. Age, obesity, pre-existing comorbid illness, poor diet and sedentary lifestyle leading to lower sun exposure and lower oral intake of vitamin D are all potential confounders of any association. These confounders are difficult to fully control for in observational studies, and while large randomized controlled trials of vitamin D supplementation specifically for treating cardiovascular disease are underway the results have not been reported.

Mendelian randomization is a method where measured variation, such as single nucleotide polymorphisms (SNP), in genes directly influencing a modifiable exposure, are used to examine the causal relationship between a modifiable exposure and a disease[332]. This method provides an opportunity to overcome some of the challenges of confounding encountered in observational studies[333] and has proved valuable in determining the relationship between cardiovascular disease and other associated factors such as the interleukin-6[168], C-reactive protein[334], and triglycerides[335].

Recent data show that single nucleotide polymorphisms located within or near 4 genes related to vitamin D metabolism: rs2282679 in vitamin D binding protein ; rs12785878 in 7-dehydrocholesterol reductase (*DHCR7*); rs10741657 in cytochrome P450 2R1 (*CYP2R1*) and rs6013897 in *CYP24A1* are associated with differences in 25OHD levels[2, 6]. If the relationship between 25OHD levels and CAD is causal then we would expect 25OHD decreasing alleles to increase the risk of coronary artery disease to the same extent as their influence on decreasing circulating 25OHD[332].

Genotyping information from these 4 genes was combined with phenotypic and long-term outcome data from a large, well-characterized group of patients. This information was analysed

using a Mendelian randomization study design to investigate the causal relationship between 25OHD levels and coronary artery disease events.

6.2 Methods

6.2.1 25-hydroxyvitamin D levels

Serum 25OHD levels were measured using tandem mass spectroscopy at a single centre (Dept. of Clinical Chemistry, Glasgow Royal Infirmary, and Glasgow, Scotland) and were available for approximately 5% of patients in this study. Where more than one measure was available for an individual a mean value was used. As 25OHD tests had been requested by clinicians for specific clinical indications rather than being a random sample we compared demographic data for this subgroup with the main study.

It is well established that 25OHD levels are greatly influenced by season so a seasonal adjustment was made to all 25OHD levels using the mean difference obtained from paired t-tests[227] in a subgroup of 14 participants who had 25OHD measures for all four seasons. The seasons were defined according to the Wang et al., 2010 criteria as spring (April-June), summer (July-September), autumn (October-December), and winter (January-March) - which was used as a reference for the seasonal adjustment[6]. A linear regression was performed where the effect of age, sex, body mass index and season of measurement were determined on a single, seasonally corrected 25OHD measure for each of the participants that had 25OHD measurements. The null hypothesis was tested that there would be no significant impact of season of measurement on seasonally adjusted levels.

6.2.2 Coronary artery disease definition

Coronary artery disease cases and controls were defined using the criteria described in chapter 3.

6.2.3 Genotyping

Rs2282679 (ABI assay ID: C__26407519_10), rs6013897 (ABI assay ID: C__29958084_10), rs10741657 (ABI assay ID: C___2958430_101) and rs12785878 (ABI assay ID: C__32063037_10) were genotyped in the Go-DARTS study using TaqMan-based allelic discrimination[336] described .

An exact Hardy-Weinberg equilibrium test was performed on the genotyping data and duplicate samples were used to test for concordance.

6.2.4 Effect of 25OHD on coronary artery disease

It was only possible to determine the effect of 25OHD on CAD status in 5% of the total population who had 25OHD levels and conformed to the case/control definition. CAD cases were defined as above and controls were defined as participants with no known cardiovascular disease (including stroke, CAD and lower extremity arterial disease). A Logistic regression using a case control analysis was required in this context as available 25OHD levels were obtained as clinically indicated and may therefore have been obtained retrospectively or prospectively with respect to study enrolment. Age, sex, history of smoking, hypertension and body mass index were included as covariates. The effect of square root transformed 25OHD levels on CAD status was determined and this value was used to calculate the expected effect of the genotype score on CAD status. To show that decreasing 25OHD levels were associated with CAD status, mean, seasonally corrected 25OHD levels were associated with CAD as categorical values of high to low 25OHD levels: high (> 40nmol/L) 25OHD, medium (>22nmol/L and <= 40 nmol/L) and low (<=22 nmol/L).

6.2.5 Association of effect alleles from established vitamin D reducing loci in the Go-DARTS study

The per allele reduction in square root transformed 25OHD measures were determined in this study and compared to the published data for the four SNPs known to be associated with reducing 25OHD levels. To calculate effect estimates compatible with published effect estimates the SNPs were associated with square root transformed seasonally corrected 25OHD using an additive model in a multivariate linear regression[227] while correcting for age, sex and body mass index. The effects of rs2282679 and rs10741657 on square root transformed 25OHD levels were compared to those observed by Ahn et al., 2010[2]. The effects of these SNPs on untransformed 25OHD measures using the same model but the 25OHD measures were not transformed.

6.2.6 Comparison of 25OHD reduction by genotype from the GoDARTS study with published data

The mean 25OHD measures for individuals that were homozygous for the 25OHD reducing alleles were compared to the mean 25OHD measures for individuals carrying no 25OHD reducing alleles (0 vs. 2) for each of the four SNPs investigated here. A percentage difference between the two groups was calculated and this was compared to the percentage change observed for these two groups by Wang et al., 2008 and Ahn et al, 2010[2, 6].

6.2.7 Genotype score calculation

Genotype scores were calculated for each participant in the study. The genotypes of the four SNPs were coded as 0 - no effect alleles present; 1 - one effect allele present and 2 - two effect alleles present. Effect alleles were defined as those alleles reported to decrease 25OHD by Wang et al., (2010) and Ahn et al., 2010[2, 6] and were also confirmed in the present study. A genotype score (GS) was calculated using the methodology described under heading 2.9.

6.2.7.1 Genotype score 1

Allele effect estimates derived from large meta-analyses were preferred to those obtained from a single study as these are more likely to have greater precision as estimates of the actual allelic effect sizes. Meta-analysis effect sizes were calculated for 25OHD measures on the square root scale so the effect estimates for each additional effect allele on the square root of mean, seasonally corrected 25OHD was estimated using a linear regression while correcting for mean age, sex and body mass index. The allele effect estimates for rs2282679 and rs10741657 (a known proxy for rs2060793, r^2=1) on square root transformed 25OHD levels were published[2]. Although rs6013897 and rs12785878 had been reported to decrease 25OHD levels the meta-analysis allele effect estimates were not published[6]. Therefore a primary genotype score "GS1" was calculated by weighting rs2289679 and rs10741657 by the published allelic effect estimates combined with allelic effect estimates from this study for rs6013897 and rs12785878.

6.2.7.2 Genotype score 2

A secondary genotype score "GS2" was calculated by weighting each effect allele by the size of its effect on mean, seasonally corrected 25OHD levels. This score was completely based on

within study estimates and was used to determine the effect of per nmol/L reductions in 25OHD on the increased risk of incident CAD within this study.

Tertiles of GS1 and GS2 were determined for the entire study and were not sub analysis specific. Tertiles of GS1 were associated with 25OHD levels in a linear regression while correcting for average age, sex and body mass index. This analysis was repeated for GS2. GS1 and GS2 were tested for association with incident CAD in a survival analysis.

6.2.8 Effect of genotype score on the risk of CAD

Cox's regression analyses were performed to compare event rates for CAD across tertiles of GS1[337]. Participants with no prior stroke or CAD events were followed up from enrolment (date of genotyping) until first ever CAD event, death or end of the observation period. To demonstrate that the results were not biased by the allelic effect estimates used to calculate the genotype scores, analyses were repeated using GS2.

Analyses were performed a) unadjusted, b) adjusted for age, sex, body mass index and history of smoking, which are known to affect 25OHD levels (model 1), c) further adjusted for covariates recorded at baseline: a diagnosis of hypertension (based on a recorded blood pressure reading >160/90 mmHg or a prescription of antihypertensive medication), a diagnosis of diabetes mellitus and total cholesterol level (model 2). A final adjustment was made (model 3 – fully adjusted model) by including a prescription for statin medication at baseline. Statins were specifically included as a covariate as some statins have been reported to increase vitamin D levels[338, 339] while other studies have shown that vitamin D insufficiency may decrease the lipid lowering effect of some statins[340].

Individual SNPs were also tested for association with incident CAD in a Cox's regression with adjustment for age and sex to assess the contribution of individual SNPs to CAD risk. All analyses were performed using the R statistical package version 2.12.0[227].

6.2.9 Observed vs. expected effects of the GS1 on CAD

In order to determine if the observed effect of GS1 on risk of CAD was consistent with what would be expected based on the effect of GS1 on 25OHD levels we used a method proposed by

Freathy et al. (2008)[341]. Under the assumption that GS1 affects the risk of incident CAD through its influence on 25OHD levels, it is possible to determine an expected value for the effect of the genotype score on the risk of CAD as *c=a x b*, where *c* is the expected effect of the genotype score on CAD risk; *a* the observed effect of the genotype score on square root transformed 25OHD levels and *b* the observed effect of square root transformed 25OHD on CAD (Figure 1-7). Heterogeneity between the observed and expected effects of GS1 was calculated using the equations described under heading 2.12. The null hypothesis was that there was no difference between the observed and expected effect sizes of GS1 on risk of CAD. If there was no statistically significant difference between the observed and the expected effect (p=>0.05) then there was a failure to reject the null hypothesis but in cases where p<0.05 then the null hypothesis was rejected and the observed effect was not consistent with that expected. In this case the effect of GS1 would have mediated the risk of CAD through factors external to its effect on decreasing circulating 25OHD levels.

6.2.10 Power calculation

The power to detect an effect of 250HD and GS1 on CAD were calculated using the method described by Demidenko, 2007 [342].

6.3 Results

6.3.1 Study population

11,332 participants from Go-DARTS who had complete baseline data and had a genotype call rate of at least 75% were included in the study population and baseline characteristics are given in Table 6-1. For the incident CAD analysis a further 878 participants with previous CAD events were excluded leaving 10,454 participants in the incident CAD analysis. 25OHD levels were available for 599 participants and characteristics for this sub-population are given in Table 6-1. This sub-population was slightly older than the wider study population; was enriched for female individuals, and individuals with diabetes (all t-test p-value <0.05), however they were not dissimilar from the wider study population for BMI and history of smoking (t-test p-value>0.05) (Table 6-1). The median 25OHD level in this subpopulation was 43 nmol/L (IQR 23 to 58).

Table 6-1: Study population characteristics and genotype frequencies

Characteristic		25OHD	Overall	GS1 tertile 1	GS1 tertile 2	GS1 tertile 3
Male (%)		238 (40)	6023 (53)	2001 (53)	1874 (53)	1909 (54)
Mean age at genotyping (SD)		68.7 (12)	63.4 (13)	63.2 (13)	63.5 (12)	63.7 (12)
Diabetes mellitus (%)		351 (59)	5661 [52]	1849 (49)	1780 [52]	1790 [52]
Body mass index (SD)		28.2 [168]	29.3 [168]	29.3 [168]	29.4 [168]	29.3 [168]
History of smoking (%)		341 (57)	6653 (59)	2205 (59)	2054 (58)	2124 (60)
rs2282679 (G)	0	280 (51)	5595 (49)			
	1	225 (41)	4737 (42)			
	2	39 (8)	1000 (9)			
rs6013897 (T)	0	359 (68)	7438 (66)			
	1	151 (29)	3474 (31)			
	2	18 (3)	420 (4)			
rs10741657 (G)	0	80 (17)	1837 (16)			
	1	249 (51)	5463 (48)			
	2	158 (32)	4032 (36)			
rs12785878 (G)	0	371(68)	7563 (67)			
	1	163(30)	3351 (30)			
	2	12 (2)	418 (4)			

6.3.2 Seasonal correction of 25-hydroxy vitamin D measures

For single 25-hydroxy vitamin D measures per participant (N=599) that were not adjusted for season of measurement the season of measurement showed a reduction of -0.09nmol/L (Standard Deviation=0.02) of circulating 25OHD for every season change from Spring to Winter p=9.8E-05. The linear regression was corrected for age, sex and body mass index. The paired t-tests of intra-individual variation showed that there was a substantial difference when comparing the mean 25OHD measure for winter with spring, summer and autumn individually (N=14). The mean difference between winter and spring was -3.37nmol/L, winter and summer was -9.12nmol/L, and winter and autumn was -0.00017nmol/L. Study-wide measurements of 25OHD were adjusted by subtracting the mean difference from the measured 25OHD according to the season the test was requested in, while winter measurements were unadjusted. Using the same single 25OHD measures as before the season was associated with the seasonally corrected 25OHD measure while correcting for age, sex and body mass index. The season of measurement was no longer associated with the seasonally corrected 25OHD measures p=0.54 and the seasonal effect was greatly reduced -0.01nmol/L (Standard Deviation=0.02) per season change from spring to winter.

6.3.3 Genotyping

Rs2282679 (MAF=0.26, P_{HWE} =0.992, percentage genotyped = 99.5%, concordance=0.99), rs6013897 (MAF=0.19, P_{HWE}=0.4, percentage genotyped = 97%, concordance=1), rs10741657 (MAF=0.40, P_{HWE} =0.37, percentage genotyped=90%, concordance=1), rs12785878 (MAF=0.18, P_{HWE} =0.66, percentage genotyped = 99.7%, concordance=1) were successfully genotyped in the Go-DARTS study.

6.3.4 Effect of 25OHD on coronary artery disease

Square root transformed 25OHD levels were associated with a reduced risk of CAD, OR=0.80, 95% CI (0.66, 0.99), p=0.034 (Table 6-2). This corresponded to a beta coefficient effect of -0.21, SE=0.10, p=0.039. This showed that as the square root transformed value increased so the risk of CAD decreased. When decreasing 25OHD levels were associated with CAD as categorical variable ranging from high to low, the category representing 25OHD levels in the range of 22nmol/L to 40nmol/L associated with increased risk of CAD OR=1.81, 95% CI (0.76 to 4.3), p=0.18 and the lowest category representing 25OHD measures of less

than 22nmol/L associated with an increased risk of CAD OR=3.94, 95%CI (1.62 to 9.55), p=2.45E-03 (Table 6-2).

Table 6-2: Association of square root transformed 25OHD measures and categorical measures of 25OHD grouped into high, medium and low 25OHD levels with coronary artery disease.

Model	Covariate	OR	95%CI	P value
1	Square root transformed 25OHD	0.80	(0.66-0.99)	0.034
2	High 25-OHD (>40)	1	-	Ref
	Medium 25-OHD (>22 and <= 40)	1.81	(0.76,4.3)	0.180
	Low 25-OH D (<22 nmol/L)	3.94	(1.62,9.55)	0.002

Model 1: Association of square root transformed 25OHD levels with outcome. The model is corrected for age, sex, hypertension, smoking history and body mass index.

Model 2: Categories of 25OHD levels medium and low are compared to the highest tertile. The model was corrected for age, sex, smoking history, hypertension and BMI.

6.3.5 Effects of 25OHD reducing alleles on square root transformed 25OHD levels

The effect of the 25OHD reducing allele of rs2282679 on square root transformed corrected 25OHD levels was -0.46 (SE=0.15), p = 1.6E-03 (Table 6-3). Compared to the meta-analysis estimate of -0.36 (SE=0.05), p=1.1E-22[2] both estimates have overlapping confidence intervals. The effect of the 25OHD reducing allele of rs10741657 on square root transformed 25OHD levels was -0.10 (SE=0.13), p=4.6E-01. Compared to the published effect of -0.25 (SE=0.05), p=2.9E-17[2] these two estimates show overlapping confidence intervals. Rs6013897 showed a per minor allele effect of -0.21 (0.16), p=1.8E-01, on square root transformed 25OHD (Table 6-3). The minor allele of rs12785878 showed a per allele reduction in square root transformed 25OHD of -0.17 (SE=0.17), p=3.0E-01 (Table 6-3). Comparable values for these two SNPs were not presented in the literature.

Table 6-3: Association of SNPs previously associated with vitamin D insufficiency with square root transformed 25 hydroxyvitamin D levels in a subgroup of the main study (N=599)

SNP ID	Effect allele	AF*	Effect	95% CI†	P value	Gene
rs2282679	C	0.278	-0.46	(-0.75,-0.18)	1.6E-03	GC‡
rs6013897	T	0.176	-0.21	(-0.54,0.10)	18E-01	near CYP24A1§
rs10741657	G	0.569	-0.10	(-0.36,0.17)	4.7E-01	near CYP2R1‖
rs12785878	G	0.170	-0.17	(-0.51,0.16)	3.0E-01	DHCR7#
Genotype score			-0.98	(-1.54,-0.43)	5.2E-04	

Genotype Score	Effect on 25OHD (nmol/L)	95%CI	P value
GS1	-16.08	(-25.30,-6.87)	6.5E-04

*Allele frequency; †Confidence interval; ‡Vitamin D binding protein; §1,25-dihydroxyvitamin D (3) 24-hydroxylase; ‖ Vitamin D 25-hydroxylase; #7-dehydrocholesterol reductase

6.3.6 Effects of 25OHD reducing alleles on untransformed 25OHD levels

The minor allele of rs2282679 showed a per allele reduction in 25OHD of -6.44nmol/L, 95%CI (-10.32,-2.57), p=1.6E-03 per minor allele in the Go-DARTS study (Table 6-4). The major allele of rs10741657 showed a per allele reduction in 25OHD of -0.94nmol/L, 95%CI (-4.49, 2.60), p=6.1E-01 (Table 6-4). The minor allele of rs12785878 showed a per allele reduction of -2.01nmol/L, 95%CI (-6.49, 2.47), p=3.8E-01 (Table 4). The minor allele of rs6013897 showed a per allele reduction in 25OHD of -2.4 nmol/L, 95%CI (-6.68, 1.88), p=2.7E-01 (Table 6-4).

Table 6-4: Association of 25OHD lowering SNPs with mean seasonally corrected 25OHD measures in the Go-DARTS study

SNP ID	Effect allele	AF*	Effect on 25OHD nmol/L	95% CI†	P value	Gene
rs2282679	C	0.278	-6.44	(-10.32,-2.57)	1.6E-03	GC‡
rs6013897	T	0.176	-2.40	(-6.68,1.88)	2.7E-01	near CYP24A1§
rs10741657	G	0.569	-0.94	(-4.49,2.60)	6.1E-01	near CYP2R1||
rs12785878	G	0.170	-2.01	(-6.49,2.47)	3.8E-01	DHCR7#
GS2			-1.00	(-1.54, -0.47)	2.7E-04	

*Allele frequency; †Confidence interval; ‡Vitamin D binding protein; §1,25-dihydroxyvitamin D (3) 24-hydroxylase; || Vitamin D 25-hydroxylase; #7-dehydrocholesterol reductase

6.3.7 Comparison of difference in mean 25OHD levels between genotype groups in GoDARTS compared to published data

The mean 25OHD measure for rs2282679 in individuals who had no 25OHD reducing alleles (Group 0) was 42.7nmol/L (SD=26.7nmol/L) and in the group that were homozygous for the 25OHD reducing alleles (Group 2) was 31.5nmol/L (SD=21.9nmol/L), this corresponded to a 26% reduction in mean 25OHD between the two groups. This estimate was within the percentage range of mean 25OHD reduction observed in other studies: Framingham Heart Study (FHS) of 22%, 15% reduction seen in the 1958 British Birth Cohort (1958 BC)[6] and the range of reduction observed by Ahn et al,(2010) of between -6.64 to -34.4%[2].

For rs10741657 the mean 25OHD levels for group 0 was 41.1nmol/L (23.3nmol/L) and for group 2 was 39.4nmol/L (25.96nmol/L) corresponding to a percentage reduction of 4%. The percentage difference observed in this study was within the range of percentage reduction observed in other studies: in the FHS the percentage reduction observed was 8%, the 1958 BC 7% and a range of reduction observed by Ahn et al, 2010 of between 1.5% and 14.4%.

For rs12785878 the mean 25OHD measure was 40.2nmol/L (SD=25.3nmol/L) for group 1 and 35.9nmol/L (SD=22.2nmol/L) for group 2, this corresponded to a percentage reduction of 11%. Rs12785878 was only reported as associated with 25OHD reduction by Wang et al, (2008) and the mean percentage change observed in the FHS was 10% and in the 1958 BC was 7%. Although the estimate did not fall in that range it was not dissimilar to the percentage difference observed in the FHS. No published effects on mean 25OHD by genotype were available for rs6013897.

6.3.8 25OHD Genotype scores calculated from internal and external data.

6.3.8.1 Genotype score 1

GS1 was computed, from the individual allele effects on square root transformed, seasonally corrected 25OHD given in Table 6-3, by combining effect estimates for rs6013879 (Effect= -0.21, SE=0.16) and rs12785878 (Effect=-0.17, SE=0.17) from our study with published effect estimates for rs2282679 (Effect = -0.36, SE=0.05) and rs10741657 (Effect = -0.25, SE=0.05)[2] .

GS1 had a range of 0 to 1.81; tertiles were <=0.50, >0.50 to 0.78 and >0.78. When the 599 participants in whom 25OHD levels were measured were assigned to predefined tertile

categories based on GS1, the mean 25OHD level for the first tertile was 43.3 nmol/L (SD=26.3), 38.4 nmol/L (SD=24.9) for the second tertile and 34.8 nmol/L (SD=22.6)for the third tertile (Table 3). GS1 associated with decreased 25OHD, effect= -16.08nmol\L, SE=4.69, per step increase in of GS1 and p=6.5E-04 (Table 6-3).

6.3.8.2 Genotype score 2

GS2 was calculated by using within studies estimates for a per allele reduction in mean 25OHD measures given in Table 6-4. GS2 was also associated with a 1nmol/L, 95% CI (-1.54,-0.47), reduction in 25OHD levels per step of the genotype score, p=2.7E-04, results are given in Table 6-4. The range of GS2 was 0 to 22.64; tertiles were <=3.34, >3.34 and <=8.32, and >8.32. The mean 25OHD level for tertile 1 was 44.7nmol/L (SD=27.1nmol/L), for tertile 2 was 40.5nmol/L (SD=25.4nmol/L) and for tertile 3 was 34.0nmol/L (SD=22.8nmol/L).

6.3.9 Association of 25OHD gene scores with incident coronary artery disease events

During a mean follow up of 4.5 years, a total of 254 incident CAD events occurred in 10,454 participants included in the study population.

6.3.9.1 Association of GS1 with risk of Incident CAD

GS1 was associated with an increased risk of CAD in an unadjusted model, HR=1.50 (1.07 to 2.11), p=0.020 and this finding remained significant after adjustment for multiple factors known to influence 25OHD levels and cardiovascular risk, including statin treatment (model 3), HR=1.49 (1.61 to 2.10), p=0.022 (Table 6-5). This association was most marked in the third tertile of GS1 (representing the lowest 25OHD), that showed an association with increased risk of CAD when compared to the reference tertile in an unadjusted model, HR=1.41 (1.07 to 1.86), p=0.014 and this finding remained significant after adjustment for multiple factors known to influence 25OHD levels and cardiovascular risk, including statin treatment (model 3): HR=1.39 (1.06 to 1.84) and p=0.019 (Table 6-5).

6.3.10 Power Calculations

We had 19% power to detect an effect of 25OHD levels on CAD outcomes at p<0.05, given the sample size and the effect detected in this study. We had 7% power to detect an association between GS1 and CAD outcomes in the survival analysis at p<0.05.

Table 6-5: Association between coronary artery disease events (No. cases =254 and no. controls=10,200) and tertiles of GS1 associated with square root transformed 25-hydroxyvitamin D levels

Continuous 25OHD score	**HR**	**95%CI**	**P value[#]**
Unadjusted	1.50	(1.07 -2.11)[#]	0.020
Adjusted model 3	1.49	(1.06 -2.10)[#]	0.022

Phenotype/model	**Tertile 1 HR***	**Tertile 2 HR (95% CI)**	**Tertile 3 HR (95% CI)**	**P value[#]**
Unadjusted	1	1.27 (0.96- 1.70)	1.41 (1.07- 1.86)[#]	0.014
Adjusted model 1[‡]	1	1.23 (0.93- 1.65)	1.36 (1.03- 1.80)[#]	0.028
Adjusted model 2[§]	1	1.25 (0.93- 1.66)	1.39 (1.05- 1.83)[#]	0.021
Adjusted model 3[‖]	1	1.25 (0.94- 1.67)	1.39 (1.06 - 1.84)[#]	0.019

[*]Hazard ratio; [†]Coronary artery disease; [‡]Model 1: Adjusted for age, sex, smoking history and body mass index; [§]Model 2: As model 1, plus adjusted for history of hypertension, diabetes mellitus and cholesterol level; [‖]Model 3: As per model 2, plus adjusted for statin use at enrolment into study; [#]$p<0.05$

6.3.9.2 Association of GS2 with incident CAD

GS2, which was the score weighted by per effect allele nmol/L reduction in 25OHD estimated from this study, was associated with increased risk of CAD HR=1.03, 95% CI (1.00-1.06), p=0.036 and this association remained significant after adjusting for multiple factors known to influence 25OHD levels and cardiovascular risk, including statin treatment (model 3): HR=1.03 (1.00 to 1.05), p=0.044 (Table 6-6). This corresponds to a 3% increase in risk per 1nmol/L reduction in 25OHD levels for individuals in this study. The association was most marked in the third tertile that represented the greatest reduction in 25OHD levels by genotype in an unadjusted model HR=1.44, 95%CI (1.05 to 1.97), p=0.025 and in a fully adjusted model for multiple factors known to influence 25OHD levels and cardiovascular risk, including statin treatment (model 3): HR =1.41, 95%CI (1.02 to 1.93), p=0.037 (Table 6-6).

GS2 had a range of 0 to 0.64 and was associated with decreased 25OHD, effect=-0.12nmol/L (-0.20,-0.05) per tertile increase in GS2, p=9.2E-04. Both the 2nd and 3rd tertiles of GS2 were associated with increased risk of CAD in an unadjusted model, 2nd tertile: HR=1.53 (1.14 to 2.07), p=5.3E-03 and the 3rd tertile: HR=1.50 (1.09 to 2.07), p=0.01. This association remained significant after adjustment for multiple factors known to influence 25OHD levels and cardiovascular risk, including statin treatment (model 3): 2nd tertile: HR=1.48 (1.09 to 2.00), p=0.01 and the 3rd tertile: HR=1.47 (1.07 to 2.03), p=0.02.

Table 6-6: Association between coronary artery disease events (No. cases =254 and no. controls=10,200) and tertiles of GS2 associated with 25-hydroxyvitamin D levels

Predictor/model	***HR****	***95%CI***		***P value****
GS1 continuous				
Unadjusted	1.03	(1.00-1.06)#		0.036
Adjusted model 3||	1.03	(1.00-1.05) #		0.044
GS 1 tertiles	**Tertile 1 HR**	**Tertile 2 HR (95% CI)**	**Tertile 3 HR (95% CI)**	
Unadjusted	1	1.28 (0.92-1.77)	1.44 (1.05-1.97)#	0.025
Adjusted model 1‡	1	1.25 (0.90-1.72)	1.37 (1.00-1.99)	0.050
Adjusted model 2§	1	1.25 (0.90-1.73)	1.40 (1.01-1.92)#	0.040
Adjusted model 3	1	1.25 (0.90-1.73)	1.41 (1.02-1.93)#	0.037

*Hazard ratio; †Coronary artery disease; ‡Model 1: Adjusted for age, sex, smoking history and body mass index; §Model 2: As model 1, plus adjusted for history of hypertension, diabetes mellitus and cholesterol level; ||Model 3: As per model 2, plus adjusted for statin use at enrolment into study; #p<0.05

6.3.9.3 Effects of individual SNPs on CAD outcome

To assess the direction of effect of individual SNPs compared to the effects of GS1 and GS2 on outcome, Cox's regression analyses were performed using an additive model for individual SNPs (Table 6-7). The direction of effect on CAD was consistent for the 25OHD decreasing alleles across all four SNPs although no individual SNP was significant on its own.

Table 6-7: Association between individual genotype variants and risk of cardiovascular events

Phenotype/SNP*	HR (95% CI) †	P‡
Coronary artery disease events		
Rs2282679	1.06 (0.89-15)	0.537
Rs6013897	1.20 (0.99-1.47)	0.069
Rs10741657	1.18 (0.99-1.39)	0.064
Rs12785878	1.08 (0.91-1.27)	0.625

*Single nucleotide polymorphism; †Hazard ratio and confidence interval; ‡Adjusted for age and sex the effect allele in all cases is the allele associated with lower 25OHD levels

6.3.10 The observed effect of GS1 on risk of incident CAD was consistent with the effect expected based on the genotypic effects on 25OHD levels

In this study we have shown that GS1 decreases square root transformed, seasonally adjusted 25OHD levels by 0.98 (0.25) per step of GS1 (Table 6-3) which is the value for ***a***. The results of the logistic regression demonstrated that square root transformed 25OHD were associated with decreased risk of CAD, OR=0.80 (0.66-0.99) and p=0.034 (Table 6-2) corresponding to the value for ***b***, a beta estimate of -0.21 (0.10), p=0.039 (Table 8).Thus, ***c*** = ***a*** x ***b***=-0.98 x -0.21=0.21, SE=0.18 (Table 8). This represented a hazard ratio of 1.23 (0.86, 1.75) for the expected effect of GS1 on risk of CAD while the observed estimate was a hazard ratio of 1.49 (1.06 to 2.10) (Table 6-5). Z was calculated as: Z=0.40-0.21/sqrt[0.17^2+0.18^2)], where 0.40=log(1.49) was the observed effect estimate for GS1 on risk of incident CAD , 0.21 was the expected effect estimate and 0.17 and 0.18 were the corresponding standard errors of these estimates (Table 6-8). The z-score corresponded to a p value of 0.42 indicating that there was no major difference between the expected and the observed effects of GS1 on CAD risk.

Table 6-8: The observed association between coronary artery disease and the genotype score for decreasing 25 hydroxyvitamin D levels is not significantly different from that expected based on the association of the score with 25OHD levels and the association between 25OHD levels and CAD

Covariate	Outcome	Observed effect (SE)	Expected effect (SE)	P value
25OHD levels*	CAD	-0.21 (0.10)	-	-
Genotype score	25OHD levels*	-0.98 (0.25)	-	-
Genotype score	CAD	0.40 (0.17)	0.21(0.18)	0.42

*These represent square root transformed 25OHD levels

6.4 Discussion

Mendelian randomization is the random assortment of genes that occurs as a result of meiosis. Since this process is random, genes found to modify an exposure may be associated with a disease if the exposure is causally related to the disease and this association is not susceptible to reverse causation or confounding[332]. Our results show that a composite genotype score, based on SNP alleles previously associated with lower 25OHD levels, predicted higher rates of future coronary artery disease events and the association was consistent with the effects of low 25OHD increasing the risk of CAD.

The association of the genotype score with incident CAD was not driven by a single locus and was not abrogated by adjustment for traditional risk factors. The observed effect of the genotype score on the risk of incident CAD was not different to the expected effect, which was predicted based on the effect of the genotype score on square root transformed vitamin D levels and the effect of square root transformed vitamin D levels on the risk of CAD.

It is important to note that the genotype score used in this study comprised of four SNPs localized within or near genes that have well established and specific roles in Vitamin D

homeostasis (comprising three key 25OHD metabolic enzymes and a major vitamin D binding globulin) and the increase in risk of CAD attributed to the genotype score was consistent with the increase in the risk of CAD attributed to reduced vitamin D levels. Both of these features strengthen the case that the observed results were due to vitamin D mediated effects. The 25OHD decreasing alleles for the four loci were associated with an increased risk of incident CAD and were in accordance with the overall result for the genotype score, suggesting that the results were not driven by a single locus. The effect of the genotype score on the risk of incident CAD was only marginally attenuated with adjustment for a range of factors known to be associated with lower 25OHD levels and with cardiovascular risk. The evidence presented here strongly argues for a direct role of 25OHD on influencing cardiovascular risk.

Prospective observational studies have demonstrated a relationship between lower 25OHD levels and incident cardiovascular events[326, 330, 331], in a community-dwelling population with no previous history of cardiovascular events, in patients with diabetes mellitus, and in a large group of patients referred for coronary angiography. In this study we have replicated the association of lower 25OHD levels with increased risk of coronary artery disease. However, vitamin D supplementation trials have found conflicting evidence for the association of vitamin D supplementation with cardiovascular disease event reduction.

Previously reported intervention studies have not been designed specifically to test the effect of vitamin D supplementation on cardiovascular event rates. One meta-analysis of trials conducted for other indications (mostly falls and osteoporosis) showed a 7% reduction in all-cause mortality with vitamin D supplementation[197]; it was not possible in this meta-analysis to dissect out causes of death. A more recent meta-analysis failed to confirm a reduction in all-cause mortality in patients receiving vitamin D supplementation in randomized controlled trials; this analysis also found no reduction in stroke or myocardial infarction with supplementation[343]. One large study of vitamin D supplementation reported data for cardiovascular deaths[344] and found that those receiving vitamin D had an age-adjusted relative risk of death compared to the placebo group of 0.84 (95% CI 0.65 to 1.10, p=0.2). Inadequate dose of vitamin D and co-administration of calcium may increase the risk of myocardial infarction[345], and the fact that existing trials were not designed to detect cardiovascular endpoints, may explain the lack of a convincing signal of efficacy in

intervention studies to date. In our study we showed that a 1nmol/L reduction in vitamin D corresponded to a 3% increase of incident CAD. This is larger than an estimate of an increase of 5% in cardiovascular disease and all cause death per 10nmol/L reduction in 25OHD[346]. It is possible that exposure to lower levels of vitamin D has a cumulative effect on vascular health over many decades; the fact that participants in our study had differential exposure to vitamin D from birth, courtesy of their genetic makeup and geographic location, may explain the stronger effect seen in contrast to relatively short-term intervention trials. Remarkably the association of the genotype score with incident CAD was not affected by correcting for traditional risk factors.

Little change to the strength of the associations was found after adjusting for history of hypertension, total cholesterol levels or statin treatment, suggesting that any effect of the composite genotype score and hence 25OHD may be mediated via alternative mechanisms. Vitamin D has anti-inflammatory properties[347], can improve endothelial function independently of its effect on blood pressure[348], and has a role in modifying the process of vascular calcification[349]. It is therefore possible that these mechanisms contribute to our findings.

Strengths of our study include the use of a well-characterized population with defined endpoints, enriched for people with diabetes mellitus, with a long duration of follow-up. We also used the published effect estimates for rs2282679 and rs10741657 that were based on a large meta-analysis[2] indicating that the association of the genotype score is not due to Go-DARTS study specific effect estimates. However, using a genotype score solely based on within study effect estimates of per allele nmol/L reduction in 25OHD was still associated with increased CAD risk. We also found that the association between the genotype score, that included published effects, for 25OHD decreasing alleles and CAD outcomes that was observed in this study did not differ from the association that was expected indicating that the association of the score with increased risk of CAD is consistent with the effects of the SNP alleles on lowering 25OHD levels.

Weaknesses of our study are that the number of participants with 25OHD measures and CAD events in the Go-DARTS study place limits on the statistical significance of our findings; larger studies involving combinations of existing studies will be necessary to confirm our

findings. In addition we were underpowered to detect an association between 25OHD and CAD, and an association between GS1 and CAD outcomes in the survival analysis. Despite the lack of power we do detect associations but these will need to be confirmed in larger studies. We were unable to use 25OHD levels from a random sample of the Go-DARTS population, and thus selection bias for these values is still possible, however when the mean 25OHD measures were calculated by genotype and the percentage difference between those carrying two vitamin D reducing alleles and non-carrier individuals the percentage reduction observed in this study was in the range of published reductions. The association between the genotype score and incident CAD requires replication in a much larger study, the evidence presented here is the result of a hypothesis driven experiment based on an equivocal association between 25OHD and CAD, we have indirectly confirmed the association of the genotype score with incident CAD by comparing the observed association with the association expected and found no difference. We lacked sufficient statistical power to demonstrate effects on cardiovascular event rates at the level of individual SNPs, but this was ameliorated by the greater power of the genotype score. The Go-DARTS population was overwhelmingly Caucasian in ancestry, which limits the generalizability of both the genotypic and phenotypic findings. We have chosen to focus on macrovascular outcomes in this study, as the observational evidence is strongest for the association of 25OHD levels and these outcomes. A similar examination of the relationship between 25OHD genotype and microvascular outcomes in patients with diabetes mellitus would be a target for future research.

Whilst Mendelian randomization experiments cannot substitute for randomized controlled trials, our results lend weight to a causal, albeit complex, relationship between vitamin D levels and cardiovascular disease. As such, these results strengthen the case for conducting large randomized controlled trials of vitamin D intervention.

Chapter 7: A meta-analysis of SNP effects on lower extremity arterial disease in patients with and without diabetes mellitus, and in smokers and non-smokers

Natalie van Zuydam – was responsible for the study design; writing of analysis documents for individual studies; writing up meta-analysis protocols for the meta-analysis; data procurement; the analysis of the GoDARTS data and for running the meta-analysis

Summary statistics were provided by analysts from: DCCT/EDIC, deCODE and GoDARTS

Prof Colin NA Palmer and Prof Helen Colhoun were responsible for the study design and data procurement and were senior investigators on this project

The SUMMIT proportion of this study was supported by: Mark McCarthy, Helen Colhoun and Leif Groop

7.1 Introduction

The most common method of diagnosing lower extremity arterial disease (LEAD) is an abnormal ankle brachial index (ABI) measurement of less than 0.9 or greater than 1.3. The heritability of ABI has been estimated from family studies to be between 21 and 30%[158]. To date two large-scale genetic studies have been conducted to identify the genetic determinants of LEAD that have reported rs10757278 in the 9p21 region[160] and rs1051730 in *CHRNA3*[159]. Signals in the 9p21 region have been associated with many different atherosclerotic diseases[38, 161, 273] but it remains unclear exactly what role the gene plays in atherogenesis. *CHRNA3* is an established locus for nicotine dependence and signals in the locus have been associated with increased smoking frequency[159]. Other variants in *AGT, CSMD1, ITGB3, IL6, ENPP1, OSBPL10, NOS3, MTHFR, VSP13D* and *MMP3* have been reported at nominal significance in candidate gene studies[169, 170, 176, 179, 350] but the genetic determinants of LEAD remain largely unidentified.

Smoking status is the largest predictor of LEAD that increases the occurrence of LEAD up to 10 fold[351]. Compared to other cardiovascular diseases smoking is more common in individuals with LEAD, where individuals with LEAD smoke higher quantities of cigarettes[352]. Indeed one of the genome wide association hits for LEAD is within a nicotine receptor locus *CHRNA3/CHRNA5.* Rs1051730 is associated with increased smoking quantity and has also been associated with lung cancer and LEAD through its effect on nicotine addiction[159]. However, individuals who have never smoked are still at risk of LEAD.

Diabetes mellitus increases the risk of LEAD up to three fold[353, 354] and is the second largest predictor of LEAD along with previous cardiovascular disease[353]. The prevalence of LEAD in diabetics has been estimated at above 20% compared to a much lower prevalence in non-diabetics of 12.5%[355, 356]. Prevalence and risk of LEAD are greatly increased with increasing glycated haemoglobin (HbA1c)[156, 357, 358]. The increase in risk of LEAD in diabetics can be attributed to the common risk factors for cardiovascular disease such as smoking, elevated blood pressure and dyslipidaemia[355]. Diabetics also have a higher level of vascular inflammation which puts them at higher risk of atherosclerotic disease[156].

In this study we had two main aims: to identify risk variants for LEAD and to identify variants with heterogeneous allelic effects for LEAD in smokers and LEAD in non-smokers. We

combined meta-analysis results from 4 cohorts to identify common variants associated with LEAD. We compared the findings from GWAS and candidate gene studies with the results of the meta-analysis to confirm or disprove previous findings. Finally we conducted an interaction analysis in three of the cohorts, which had smoking status available, to identify loci that had different allelic effects in smokers vs. non-smokers.

7.2 Materials and Methods

7.2.1 Study Design Summary

We combined summary statistics from 4 genome wide association studies imputed to HapMap2 in a meta-analysis. We stratified analyses in 3 cohorts by smoking status and tested loci for an interaction with smoking status.

7.2.2 Phenotype

7.2.2.1 Lower extremity arterial disease cases and controls

Individuals with LEAD were identified by individual studies based on a number of criteria given in (Table 7-1). Briefly, the main diagnosis was based on an ABI of less than 0.9 or greater than 1.4, thigh through to mid-foot amputations and corrective procedures related to LEAD. Controls were identified to be free of LEAD by individual studies.

7.2.2.2 Genotypes and imputation

Operational details of the 4 GWAS studies are given in Table 7-1. The GWAS studies were typed on the Affymetrix 6.0, Illumina Human Omni1-Quad and the Illumina Omni Express SNP arrays (Table 7-1). To obtain an increased marker set that was comparable across all GWAS platforms individual studies imputed their GWAS data to the HapMap2 CEU reference panel. Individual studies applied quality control pre-imputation and the details of that QC are given in Table 7-1. Either MACH or IMPUTE2 were used to impute missing genotypes yielding a maximum of 2,619,962 SNPs (Table 7-1).

7.2.2 Statistical methods

7.2.2.1 Individual studies

The details of the analyses performed and the software used by individual studies are given in Table 7-1. Briefly analysis software that took genotype uncertainty into account was used to analyse imputed data. Imputed SNPs were tested for their association with LEAD using a

log-additive model frequentist tests adjusted for age (onset of the first event for cases or time of recruitment for controls), gender, diabetes status and centre specific covariates to account for population structure was applied to imputed SNPs. Directly typed SNPs were modelled log-additively in a logistic regression that was corrected for age, gender, diabetes status and centre specific covariates to account for population structure. The analysis was also stratified into ever and never smokers whilst correcting for age, gender, diabetes status and centre specific covariates to account for population structure.

Table 7-1: Cohort characteristics and operational details for the studies included in the meta-analysis of lower-extremity arterial disease

Study	Go-DARTS	GoDARTS-SUMMIT	deCode	DCCT/EDIC
Full name of the study	The Genetics of Diabetes Audit and Research in Tayside Scotland	The Genetics of Diabetes Audit and Research in Tayside Scotland	deCODE	Data not yet submitted
Reference	[225]	[225]		
Ethnicity	European	European	European	
Region of recruitment	Scotland	Scotland	Icelandic	
Phenotype	LEAD	LEAD	LEAD	
Phenotype definition	Full description heading 3.2.3	Full description heading 3.2.3		
Control definition	Controls free of any vascular disease	Controls free of any vascular disease	Controls without known vascular disease	
Total sample size [men/women, cases/controls]	3532 [1906/1626,713/2819]	2427 [1431/996 ,332/2095]	28456 [11153/17303, 1458/26988]	
Mean age (SD) years [cases/controls]	64.9 (10.7) [65.6(10.5)/64.7(10.8)]	64.8 (11.8) [66.8(11.3/64.5(11.8)]	50.2 (21.4) / NA	
N LEAD cases	713	332	1458	

Study	Go-DARTS	GoDARTS-SUMMIT	deCode	DCCT/EDIC
Mean BMI (SD) kg/m2 [diab/nondiab, cases/controls]	30.4 (5.5) [30.4 5.5)/NA,30.1(5.5)/30.5(5.5)]	31.1 (6.1) [30.3(6.1)/NA,31.2 (6.1)/NA]	26.7(4.4) / 26.9 (5.4)	Data not yet submitted
N Diabetic LEAD cases	713	332	251	
Duration diabetes[case/control]	7.2 (7.5) [8.9 (10.2)/6.8(6.7)]	6.1 (6.6) [8.7(8.1)/5.7(6.2)]	NA	
N current smoker LEAD cases	175	80	NA	
N ever smoker LEAD cases	385	179	886	
Genotyping centre	Sanger	Lund	deCode	
Genotyping array	Affymetrix 6.0	Illumina OmniExpress array	Illumina HumanHap300/CNV370	
Calling algorithm	CHIAMO	Birdsuite	BeadStudio	
Pre-Imputation QC - Exclusion criteria	Heterozygous samples; cryptic relatedness and population outliers	Heterozygous samples; cryptic relatedness and population outliers	NA	
Sample call rate	>=99%	>=99%	>98%	
SNP call rate	MAF>0.01, Call rate>0.9if MAF>0.05, Call rate>=0.95 if	MAF>0.01, Call rate>0.9if MAF>0.05, Call rate>=0.95 if	>96%	

Study	Go-DARTS	GoDARTS-SUMMIT	deCode	DCCT/EDIC
	MAF<0.05	MAF<0.05		
HWE	1.00E-06	1.00E-06	10-6	
Imputation software	IMPUTE2	IMPUTE2	Impute	
Reference panel	HapMap2 and WTCCC2 controls	HapMap2	Hap Map Ceph CEU v22	
Analysis software	SNPTEST	SNPTEST	SNPTEST	
Analysis model	Logistic regression	Logistic regression	Logistic regression	

7.2.3 Data cleaning steps applied before conducting the meta-analysis

Quality control of the data was performed centrally. SNPs with a MAF lower than 1% or those with less than 10 minor allele homozygotes calculated as (2 x (Number of cases) x MAF)) were removed from individual studies. Specific quality control criteria were applied to directly typed and imputed SNPs separately. Directly typed SNPs with a MAF greater than 5% were removed if the p value for the exact test for Hardy-Weinberg equilibrium (HWE) was less than 5.7E-07 and for SNPs with a MAF less than 1% with a less than 1E-04. Imputed SNPs were removed if they had an R^2 value less than 0.3 if imputed with MACH and an R^2 value less than 0.4 if imputed using IMPUTE2. If imputed SNPs were analysed using SNPTEST then imputed SNPs were removed if the proper info value was less than 0.4. Pairwise allele frequency plots were drawn for each study for imputed and directly typed data separately to check for allele frequency outliers and strand differences amongst the summary meta-data submitted for analysis.

7.2.4 Statistical model checking

The summary statistics for individual studies were checked by several procedures. These included plotting the beta distribution by MAF category (0-0.01, 0.01-0.05, 0.05-0.2 and 0.2-0.5) and this was performed separately for directly typed and imputed SNPs. A deviation of the mean beta values from zero and scattered outliers were an indication of problems with the models used or low quality SNPs. To check for population stratification quantile-quantile plots were drawn from the p values submitted by each study and lambda was calculated.

7.2.5 Fixed effects meta-analysis and smoking interaction

The primary analysis comprised of an overall meta-analysis of LEAD and one sub-analysis of ever vs. never smokers. Our default meta-analysis used a fixed-effect model with inverse variance weighting and a calculation of two homogeneity statistics: Cochran's *Q* and I^2 using GWAMA (http://www.well.ox.ac.uk/gwama/)[252]. Individual studies were corrected for genomic inflation using GWAMA and the overall meta-analysis was corrected for genomic inflation by adjusting the standard error of the estimates by the inflation factor (SE*sqrt(λ)) and recalculating the p values. When there was no indication for heterogeneity for a SNP based on a Cochran's Q p value greater than 0.01 the fixed effect model was maintained. In addition to the standard meta-analysis we aimed to identify loci with heterogeneous allelic

effects in smokers vs. non-smokers and we applied the sex differentiation test in GWAMA to identify loci of interest.

7.2.6 Random effects meta-analysis

When heterogeneity was present (Cochran's Q p value <0.01) the study with the most extreme result was excluded and the meta-analysis was repeated. If heterogeneity was still apparent, we adopted and reported a random-effects model[246] for that SNP. The random effects model was applied using MetaSoft (http://genetics.cs.ucla.edu/meta/)[246].

7.2.7 Replication of known disease associations

Two genome wide associations have been reported for LEAD in the 9p21 region for rs10757278 and in *CHRNA3* for rs1051730. Other signals in candidate genes have been reported for *ENNP1* rs1044498[350] in non-smokers, *PCSK9* rs1159114[181], *ACE* rs1799752[167] and rs4340[167], *PT* rs1799963[186], *MTTP* rs1800591[177], *FGB* rs1800591[175], *IL6* rs1800795[169] and rs2069827[359], *F12* rs1801020[174], *MTHFR* rs1801133[176], *PPARG* rs1805182 and rs1805192[182], *OSBPL10* rs1902341[170], *NOS3* rs2070744[179] and rs891512[180], *VSP13D* rs235243[170], *CSMD1* rs2554503[170], *MMP9* rs3918242[169], *AGT* rs5051[169] and rs699[169], *ITGB3* rs5918[169] and *F5* rs6025[169]. Where the index SNPs were not available a suitable proxy was used (R^2>0.80). Power calculations were performed in the same way as described under heading 5.2.4.6 to identify SNPs that which should be able to replicate in the meta-analysis given the size of this study and the published allelic effects.

7.2.8 Smoking interaction

In addition to the standard meta-analysis we aimed to identify loci with heterogeneous allelic effects in smokers vs. non-smokers and we applied the sex differentiation test in GWAMA to identify loci of interest. We also tested whether SNPs known to influence nicotine addiction showed heterogeneous allelic effects for SNPs in *CHRNB3*[195], *CHRNA5*[360], *ANAPC1*[360], *SLCO3A1*[360], *CHRNA3*[361], *CTNNA3*[362] and VSP13A[362].

7.3 Results

7.3.1 Study population

The full details of the number of cases and controls and SNPs are given in Table 7-2. Briefly a total of 2774 LEAD cases were included in the overall meta-analysis: 1904 had a history of

smoking and 614 had no record of smoking. 27177 controls were included in the meta-analysis: 17941 had a history of smoking and 8112 had no history of smoking.

Table 7-2: Operational details of actual SNP numbers and final case control numbers for the different analysis subgroups

Cohort	Model	Cases/Controls	#SNPs	Directly typed lambda	#Directly typed	Imputed lambda	#Imputed
GoDARTS	All	713/2819	2445299	1.02	710775	1.02	1734523
GoDARTS-SUMMIT	All	332/2095	2085596	1.00	570751	1.01	1514832
DCCT/EDIC	All	256/1024	2592580	1.00		1.01	1607556
deCode	All	1473/21239	2428087	1.10	290349	1.10	2109512
GoDARTS	Smokers	432/1503	2445299	1.02	708376	1.02	1675491
GoDARTS-SUMMIT	Smokers	219/1192	2085596	1.02	597731	1.03	1718553
deCode	Smokers	1253/15246	2428087	1.09	290350	1.09	2110369
GoDARTS	Non-Smokers	281/1316	2445299	1.01	294355	1.01	1619708
GoDARTS-SUMMIT	Non-Smokers	113/903	2085596	1.02	579220	1.03	1605719
deCode	Non-Smokers	220/5993	2428087	1.02	235307	1.02	1479869

7.3.2 Genotypes and imputation

2614480 SNPs were analysed in the meta-analysis after QC criteria had been applied, of these 2592552 were analysed using the fixed effects method and 21928 were analysed using the random effects model for the overall analysis. A total of 2592552 were included in the smoking interaction analysis and all were analysed using a fixed effects model.

7.3.3 Fixed effects meta-analysis

The fixed effects meta-analysis in all LEAD cases and controls did not detect any signals that reach genome wide significance (Figure 7-2) although 30 signals reaching a suggestive significance level (less than 1E-04) included the known LEAD signals: rs10757269 in *CDK2NBAS* (Table 7-3, Figure 7-3A and Figure 7-4) and rs1051730 (rs8034191, R2=0.93, D'=1.00) for *CHRNA3* (Table 7-3, Figure 7-3B and Figure 7-4). The overall lambda was equal to one after correction and did not show a great deviation from the expected distribution of p values (Figure 7-1A). Suggestive signals in biologically interesting genes were found in *HIBADH, LDB3/BMPR1A, LPHN2, GAS7, CADM2* and *KAT2B* (Table 7-3).

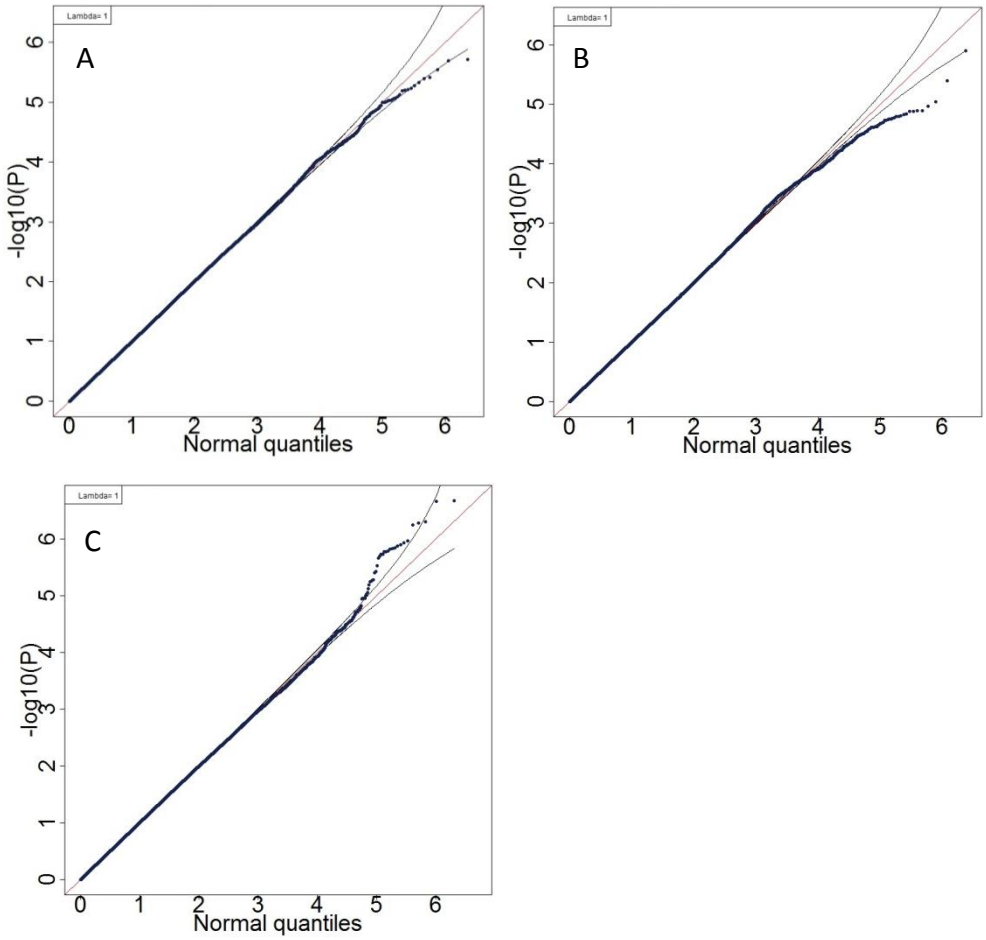

Figure 7-1: QQ plot of meta-analysis p values for all lower extremity arterial disease cases and all LEAD free controls in all individuals (A), in smokers (B) and in non-smokers (C).

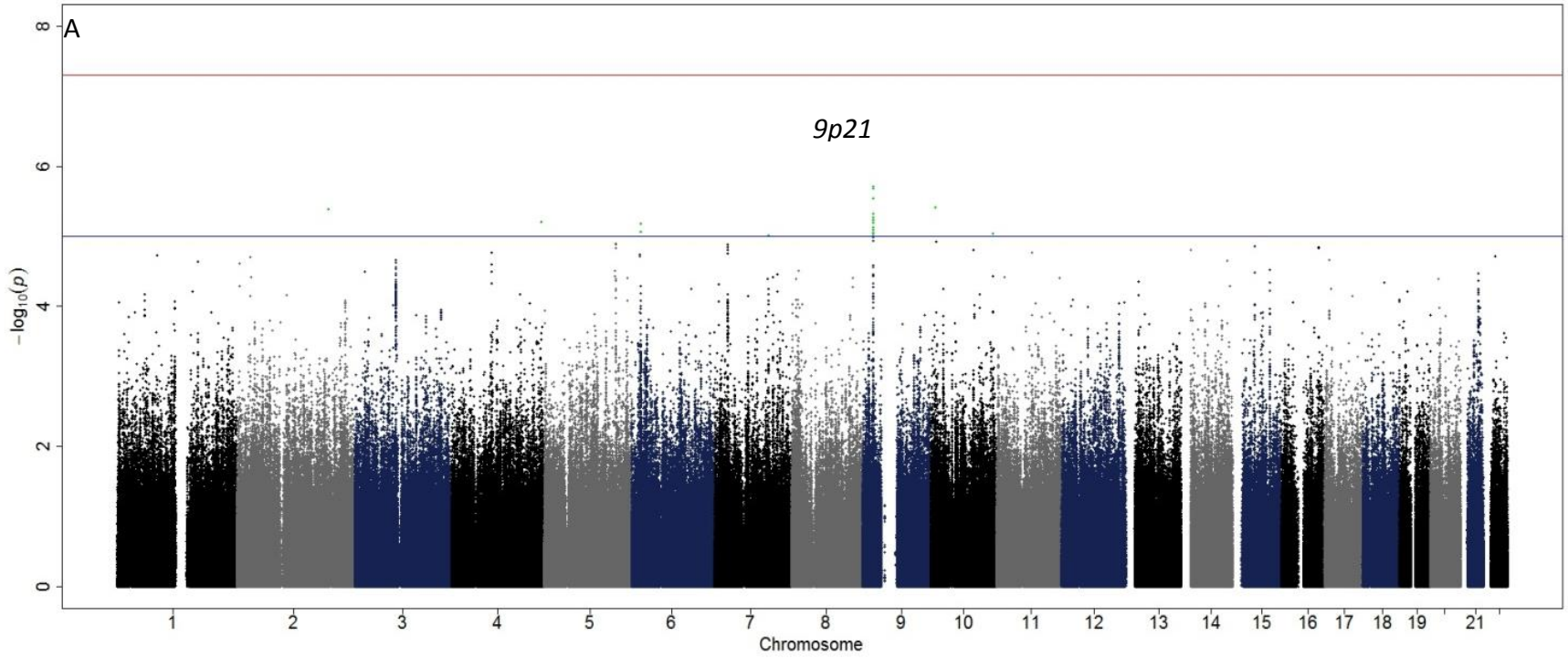

Figure 7-2: A - Manhattan plot of meta-analysis p values from all lower extremity arterial disease cases and all LEAD free controls; B - Manhattan plot of meta-analysis p values from diabetic lower extremity arterial disease cases and LEAD free controls

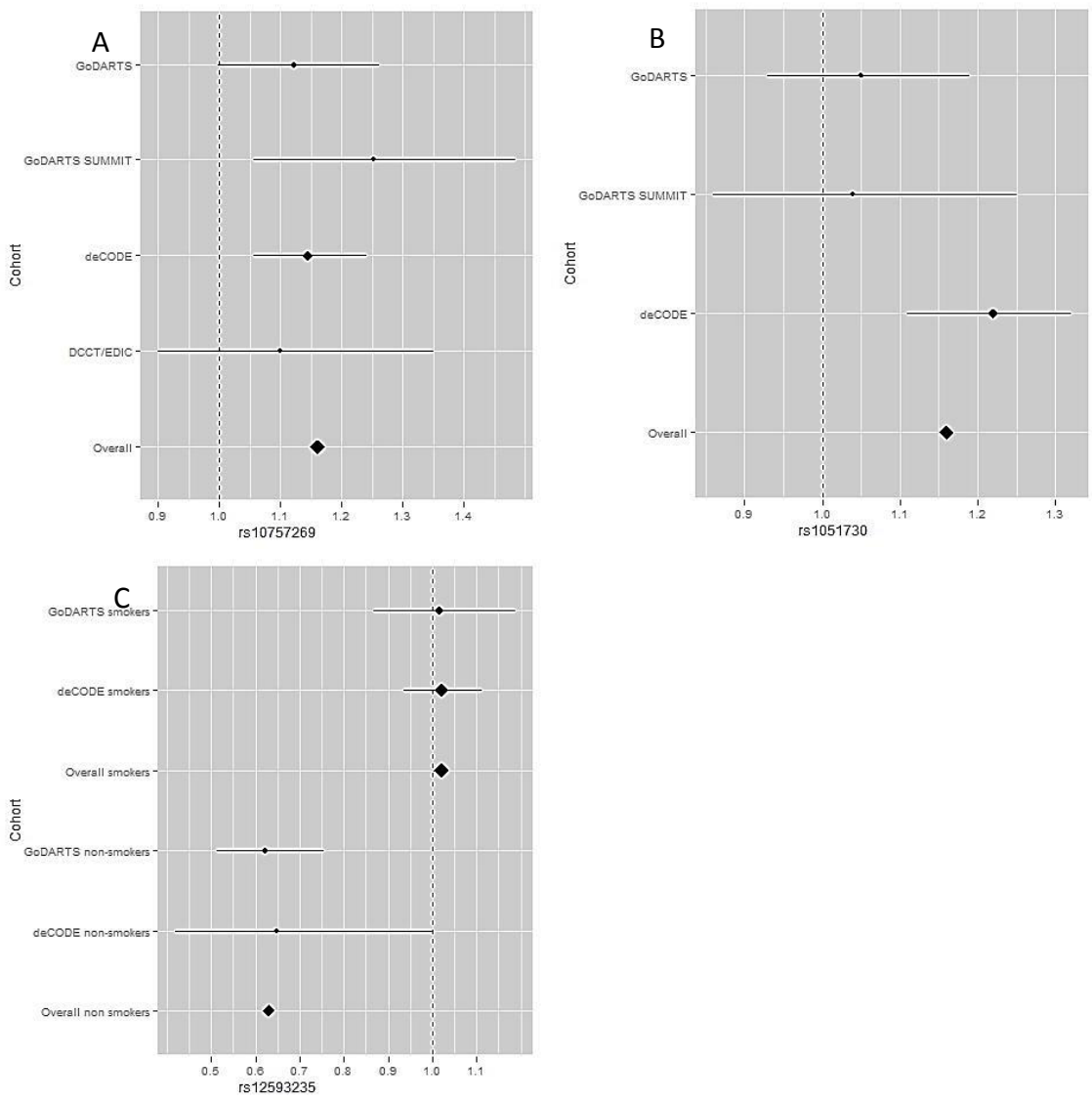

Figure 7-3: Forest plots of the 9p21 region, rs10757269 (odds ratio= 1.16), *CHRNAB3* rs1051730 (Odds ratio=1.16) and *ADAMTS17* rs12593235 in smokers (odds ratio=1.02) and non-smokers (odds ratio=0.63)

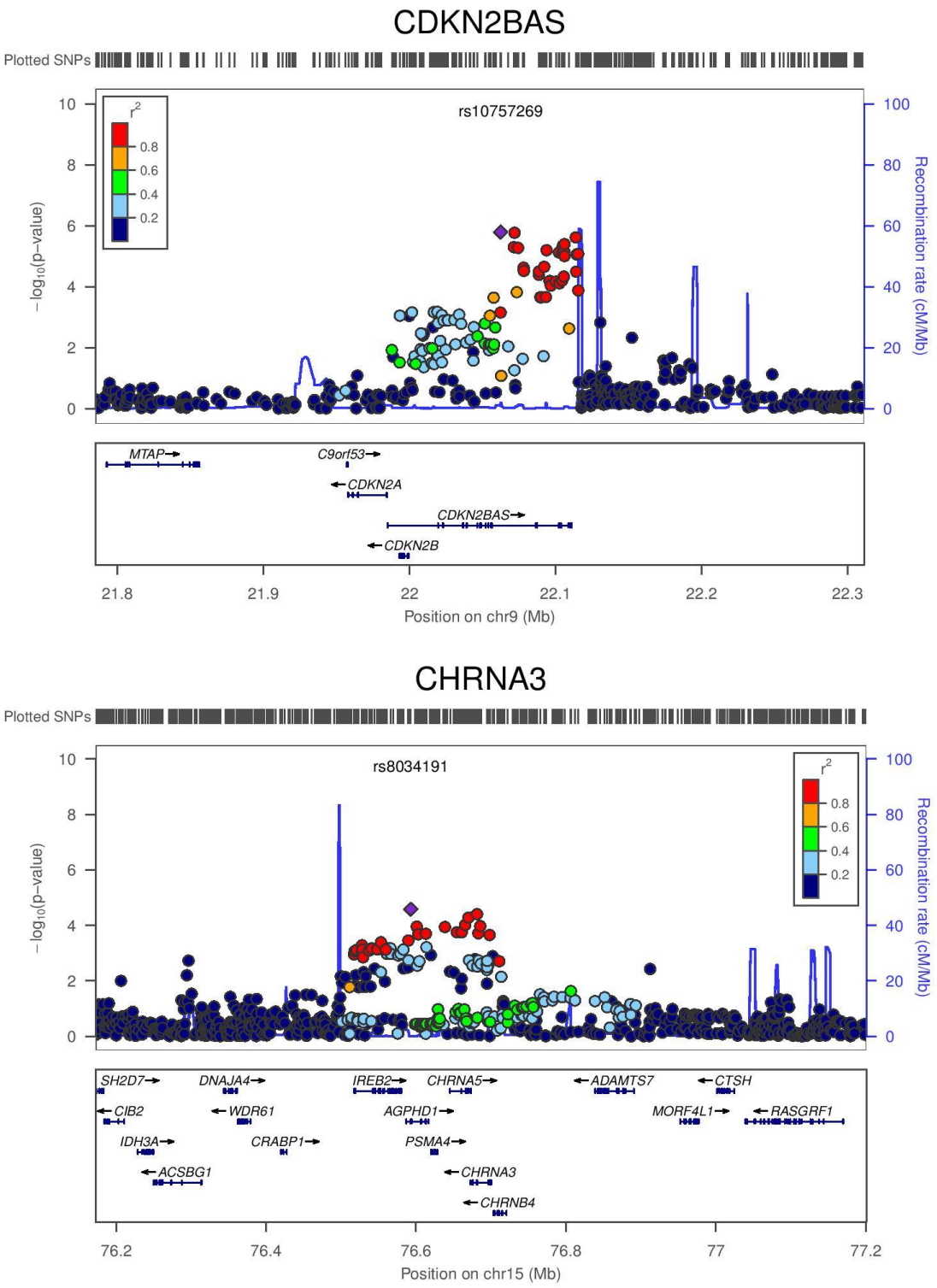

Figure 7-4: Locus zoom plots of rs10757269 in the 9p21 region which was the top hit from the meta-analysis of lower extremity arterial disease and for rs8034191 which is in linkage with rs1051730 in CHRNA3 a known LEAD locus

Table 7-3: Top hits for the overall meta-analysis for lower extremity arterial disease pruned for linkage disequilibrium

Chromosome	Position	Gene	SNPs	Effect Allele	Allele Frequency	Odds Ratio	Standard Error	P value	#Cases
9	22062264	*CDKN2B-AS1*	rs10757269	G	0.45	1.16	0.034	1.62E-06	2768
6	166560362	GAPDHP72	rs6931241	A	0.04	1.54	0.129	2.31E-06	2418
10	10328165	*CUX2P1*	rs10218946	G	0.04	2.54	0.421	3.25E-06	969
2	188259734	*CALCRL*	rs16829210	T	0.02	1.79	0.200	3.43E-06	2437
4	184744875		rs10011527	T	0.12	1.26	0.061	5.26E-06	2769
6	19022603	*TRNAQ22*/ RPL21P61	rs988458	T	0.35	1.16	0.037	5.62E-06	2507
10	129256732	*DOCK1/NPS/FOX12*	rs7088687	G	0.76	1.17	0.041	7.93E-06	2769
11	119886208	*ARHGEF12/GRIK4*	rs2244279	G	0.57	1.18	0.043	1.08E-05	2053
7	27532578	*HIBADH*	rs12700813	G	0.44	0.88	0.026	1.11E-05	2769
5	147928739	*HTR4*	rs7726693	G	0.66	0.87	0.027	1.14E-05	2769

Chromosome	Position	Gene	SNPs	Effect Allele	Allele Frequency	Odds Ratio	Standard Error	P value	#Cases
15	44936010	*SEMA6D*	rs12591095	T	0.19	1.19	0.045	1.18E-05	2747
16	77744926	*WWOX*	rs1126183	G	0.97	0.64	0.059	1.23E-05	2437
10	88374565	*WAPAL/OPN4/ LDB3/BMPR1A*	rs2355012	C	0.74	0.85	0.030	1.34E-05	2499
14	20674429	*Many genes*	rs10136872	T	0.28	1.21	0.052	1.36E-05	2511
11	73715922	*PGM2L1/P4HA3*	rs479968	G	0.40	0.87	0.027	1.46E-05	2750
4	83534645	*ENOPH1*	rs6822004	G	0.91	0.76	0.045	1.49E-05	2437
6	17169176		rs7763738	T	0.60	0.87	0.027	1.55E-05	2768
1	81912758	*LPHN2*	rs12744466	T	0.08	1.31	0.078	1.60E-05	2768
22	23753062	*KIAA1671*	rs5760804	A	0.20	1.20	0.048	1.68E-05	2162
2	28262202	*BRE*	rs7586315	T	0.07	1.34	0.087	1.72E-05	2056
17	9911846	*GAS7*	rs9913492	T	0.50	0.87	0.028	1.88E-05	2507

Chromosome	Position	Gene	SNPs	Effect Allele	Allele Frequency	Odds Ratio	Standard Error	P value	#Cases
3	85708744	*CADM2*	rs7617356	T	0.63	0.87	0.028	1.90E-05	2512
14	94430979		rs1741236	T	0.19	0.85	0.032	1.93E-05	2749
1	166436076	*TIPRL*	rs2072776	A	0.29	1.15	0.037	1.98E-05	2744
2	220553438		rs750894	G	0.77	0.85	0.031	2.39E-05	2513
15	76593078	*AGPHD1*	rs8034191	T	0.66	0.85	0.032	2.62E-05	1798
8	14852373	*SGCZ*	rs17470444	G	0.29	1.19	0.047	2.68E-05	1794
5	146651809	*STK32A*	rs7725229	T	0.15	0.83	0.036	2.74E-05	2769
3	20057973	*KAT2B*	rs2929408	C	0.87	1.24	0.059	2.77E-05	2512

*Previously associated with ankle brachial index

7.3.4 Random effects

Rs4674485 in *MIR4268* was the only SNP detected by the random effects model at a p<1E-04.

7.3.5 Replication of known signals

The top hit for the overall meta-analysis was rs10757269 that has been reported for ankle-brachial index and LEAD. We detected the same risk allele for the 9p21 region G with an odds ratio of 1.16 and the same risk allele for *CHRNA3* C with an odds ratio of 1.16 which are both similar to published estimates(Table 7-4)[169, 268]. Of the 23 SNPs previously reported for LEAD we had information for 14 of those SNPs (Table 7-4) and detected a nominally significant effect for rs5051 in *AGT* for the published risk allele[169]. A power calculation revealed that we had 100% power to detect the effects for rs1800795 in *IL6* for the published OR=1.44 for the major allele.

Table 7-4: Replication of known lower extremity arterial disease associations in the current meta-analysis for all LEAD cases and controls

Gene	SNP	Proxy	EA*	AF**	Odds ratio	Lower 95%CI	Upper 95%CI	P value	#Cases	Published Odds ratio
CDK2NBAS	rs10757278/ rs1333049	rs10757269	G	0.45	1.16	1.09	1.23	1.62E-06	2768	1.34
CHRNA3	rs1051730	rs8034191	C	0.66	1.16	1.08	1.25	2.62E-05	1798	1.19
AGT	rs5051/rs699	rs2478543	C	0.41	1.06	0.89	1.00	4.57E-02	2769	1.01
CSMD1	rs2554503	rs2554522	T	0.13	1.10	1.00	1.20	5.61E-02	2513	1.32
ITGB3	rs5918	rs8069732	C	0.87	1.08	0.84	1.01	6.83E-02	2749	0.83
IL6	rs2069827	rs2069827	G	0.91	1.09	0.82	1.02	1.12E-01	2494	Major risk allele
ENPP1	rs1044498	rs6919751	T	0.15	1.05	0.96	1.14	2.84E-01	2769	Association in non-smokers
NOS3	rs891512	rs1808593	T	0.67	1.04	0.96	1.12	3.34E-01	2765	Major risk allele
OSBPL10	rs1902341	rs1902344	C	0.40	1.03	0.97	1.09	3.55E-01	2768	1.31
IL6	rs1800795	rs7808457	A	0.55	1.03	0.91	1.04	4.10E-01	2055	1.44

Gene	SNP	Proxy	EA*	AF**	Odds ratio	Lower 95%CI	Upper 95%CI	P value	#Cases	Published Odds ratio
MTHFR	rs1801133	rs1801133	A	0.45	1.03	0.91	1.04	4.19E-01	2511	0.97
NOS3	rs2070744	rs1800783	T	0.64	1.02	0.96	1.09	4.97E-01	2766	No effect reported T ABI raising allele
VSP13D	rs235243	rs235256	A	0.08	0.97	0.87	1.09	6.32E-01	2512	1.18
MMP9	rs3918242	rs2236416	A	0.86	1.02	0.90	1.07	7.12E-01	2769	Major risk allele

*Effect allele; **Allele frequency

7.3.6 Smoking interaction analysis

Figure 7-5 includes the Manhattan plots for smokers (A) and non-smokers (B) which show different patterns of association in the two strata. Figure 7-1B shows the quantile-quantile plot for associations in smokers that appear slightly deflated and C shows a deviation outside the 95% confidence intervals in non-smokers. Two SNPs were strongly suggestive of a smoking interaction (p<1E06) near *ADAMTS17* and *UMODL1* (Figure 7-5). A forest plot of the effects in non-smokers vs. smokers for rs12593235 clearly illustrates the different allelic effects in the two strata (Figure 7-3C).

Other interesting interactions were observed for signals in *GSX1, SORBS1, ARHGAP242, PACRG, FRS3, TFEC, LINGO2, ZNF385B, MTAP/CDK2NBAS, SPC25* and *STARD13.* We did not detect an interaction with smoking signals in *CHRNB3*[195], *CHRNA5*[363], *ANAPC1*[360], *SLCO3A1*[360], *CHRNA3*[361], *CTNNA3*[362] and VSP13A[362]. A nominally insignificant interaction was observed for rs1051730 at p=0.08. The ratio of smokers to non-smokers LEAD cases was 3:1 and in the LEAD controls was 2:1, so we were underpowered to detect small differences in allelic effects between smokers and non-smokers. *ENNP1,* rs1044498, was detected in non-smokers only but we found no association between LEAD and rs1044498 in the overall meta-analysis and there was no interaction with smoking status.

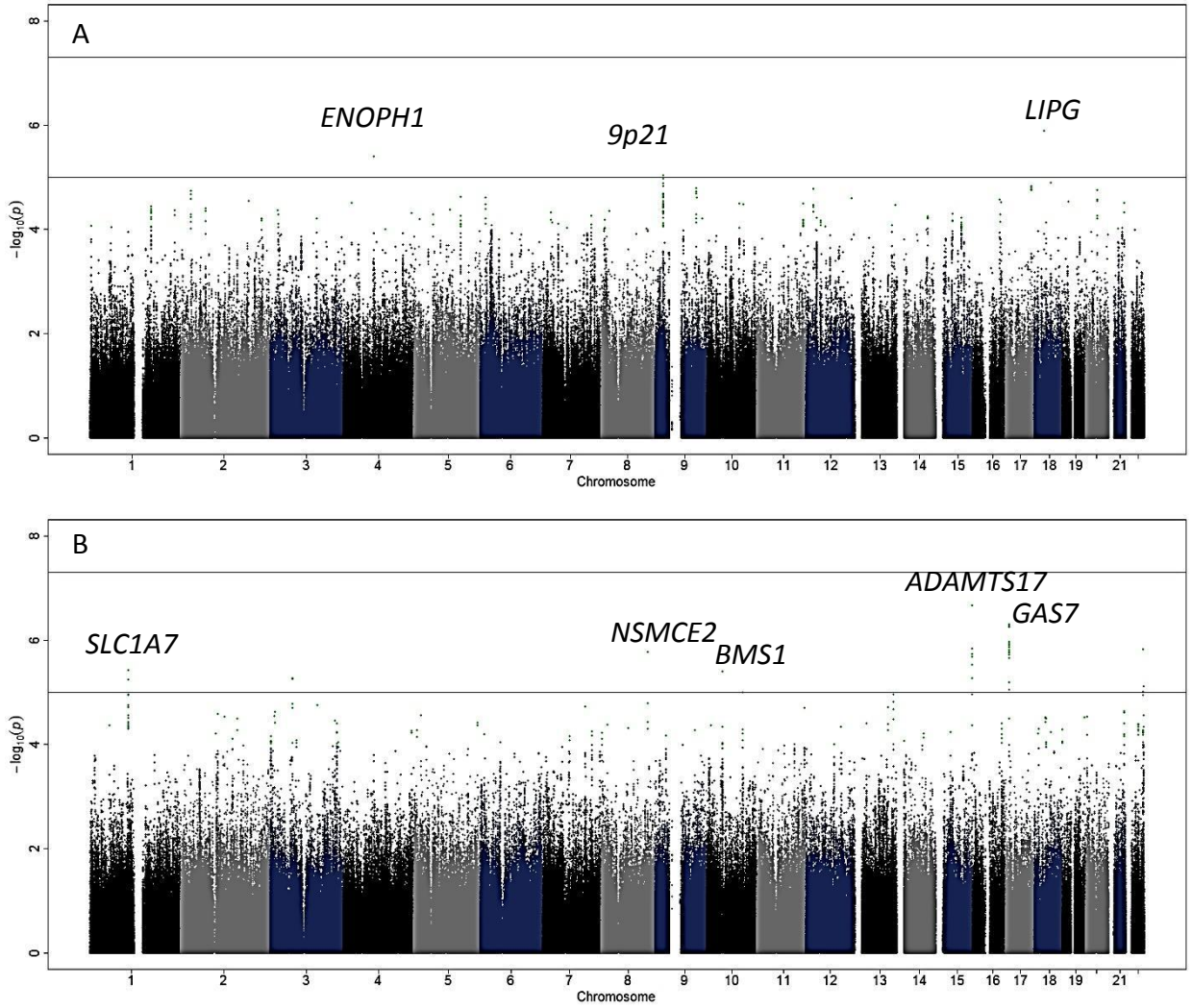

Figure 7-5: Manhattan plots of LEAD determinants in smokers (A) and non-smokers (B)

7.4 Discussion

In this study we present findings from a meta-analysis of LEAD in a total of 2774 LEAD cases and 17941 LEAD free controls and the first meta-analysis of a gene by smoking interaction for LEAD. We replicate known signals in the 9p21 region and *CHRNA3* that are the only signals reported at genome wide significance for LEAD. We also investigated the published signals from candidate gene studies and detected an association with *AGT.* Two signals near *ADAMTS17* and *UMODL1* had large differences in allelic effects between smokers and non-smokers.

The 9p21 region and *CHRNA3* have both been associated with LEAD in other studies[159, 268]. The 9p21 region is a locus that has been associated with different forms of atherosclerosis including CAD, ischaemic stroke, abdominal aortic aneurisms and intracranial aneurisms[38, 161, 273]. Loss of *CDKN2B* in mice led to apoptosis of the smooth

muscle cells leading to aneurism formation[364]. The action of genes in the 9p21 region is complex where *MTAP* has been shown to increase lesion size and knock-outs of *CDKN2A* reduce lesion size in mice[285]. *CHRNA3* increases the risk of LEAD by predisposing individuals to nicotine dependence[159].

We did not detect signals for all the variants that have been associated with LEAD in candidate gene studies. While the effect sizes and risk alleles were consistent for most loci we only detected one locus at nominal significance. *AGT* is angiotensinogen that is primarily expressed in the liver where it acts through different pathways to regulate blood pressure. Mutations in this gene increase the susceptibility to hypertension – a known risk factor for LEAD[365]. Signals in *ITGB3, MTHFR* and *VSP13D* did not have the same effect allele as previously published but none of these results are significant at $p<0.05$. We did not detect an association in *IL6* with the reported effect size despite the power to detect an association in this study. Larger studies are required to identify true genetic associations for LEAD.

Other suggestive signals from the meta-analysis were in genes that may be linked to the disease process or may be related to risk factors for LEAD. Body mass index, diabetes and increased blood sugar are known risk factors for LEAD. We detected signals close to known BMI loci in *CADM2*[366], in *HIBADH* which is upstream of *JAZF1* a known locus for T2D[367] and rs1741236 which is close a locus associated with response to statins[368]. LEAD and diabetic neuropathy are often diagnosed together so genes that are involved in signal transduction and cell to cell adhesion, cell growth and differentiation such as *LPHN2* and *KAT2B* are good candidates for the disease process. Advanced stages of LEAD are often characterised by a stiffening of the arteries so genes that are involved in osteoblast development and matrix mineralisation such as *GAS7*[369] and *BMPR1A*[370] are also plausible associations. Large studies will be required to confirm these signals.

The top association for the smoking interaction is near *ADAMTS17* a metallopeptidase with a thrombospondin type 1 motif. Signals in *ADAMTS17* have been associated with human height and short stature[371, 372]. This locus forms part of a long range quantitative trait locus which affects blood pressure specifically in non-smokers[373]. Our lead SNP rs1293235 shows a similar pattern of association where there is no allelic

effect in smokers but is 1.4 fold riskier in the non-smokers. In a study of Mexican Americans two SNPs in *ADAMTS17* were associated with blood pressure traits: rs8027190 was associated with hypertension and rs2573652 was associated with diastolic blood pressure[373]. These signals were not in linkage with our top associations but we observe a similar association to the one that has been as previously published. Replication of these findings in independent samples will be required to establish this locus as a risk locus for LEAD specifically in non-smokers. We detected another signal that showed heterogeneous allelic effects downstream of *ARHGAP42* another locus for blood pressure[31].

We did not detect any significant interactions for the nicotine associated loci and we believe this is due to the nicotine SNPs effect on partitioning smoking status. This would result in a large allele frequency difference between smokers and non-smokers. The effect that may be detected by an interaction analysis is the additional effect of the SNPs on smoking intensity within the smoking group. Taking into account the size of the study and the modest effect that these SNPs have on smoking intensity we would not expect to detect an interaction as it is likely that our study is underpowered. We do detect an interaction for the major allele of rs12593925 in *ADAMTS17* as it has a large allelic effect in the non-smokers and no effect in the smokers. We did not detect an association with *ENNP1* with LEAD in non-smokers which is the subgroup that it had been reported in[173].

Other interesting signals were detected in genes that modify risk factors and are associated with comorbidities of LEAD. Signals in *UMODL1* have been associated with non-cardio-embolic stroke[374] and myopia[375]. Signals in *GSX1* are risk in smokers and protective in non-smokers, these variants are downstream of rs9512637 a known alcoholism variant[376]. We detect rs483109 close to other signals for HbA1c[377]. HbA1c at high levels may be linked to an increase in vascular calcification[378]. Other signals were found close to signals for intracranial aneurisms for rs9315202 in STARD13[379] and myocardial infarction for rs4512473, close to *MTAP*. These are both known comorbidities and complications of LEAD.

The main weakness of the study was lack of power to detect variants at genome wide significance in both the overall GWAS and the interaction analysis. The power to detect

the known signal rs10757269 in this study at p<5E-08 was 0.03 so this study is underpowered to detect novel signals at this significance level with similar odds ratios and effect allele frequencies. The main GWAS included 2774 LEAD cases and while we replicated both the known signals for LEAD we were unable to detect either at genome wide significance. We were underpowered in the interaction analysis to detect significant interactions as the ratio of cases in the ever smoker and never smokers was 3:1. Despite this we did detect heterogeneous effects in a known smoking interactions locus.

In this study we have detected signals for known associations in *CHRNA3* and the 9p21 region, and have identified several signals that may be associated with LEAD. We also report the first locus that has interactions with smoking status for LEAD. We plan to seek replication for this locus and to refine signals detected in both the overall meta-analysis and the interaction analysis by including more samples in the meta-analysis. This study includes 3 diabetic cohorts (DCCT/EDIC, GoDARTS and GoDARTS-SUMMIT) where LEAD has been identified as a diabetic complication and one cohort that been selected as a case control study for LEAD (DeCODE). Only 30% of the deCODE cohort were diabetic and we hypothesize that there are probably two underlying aetiologies for LEAD in diabetics and LEAD in non-diabetics. Genes underlying the two distinct LEAD aetiologies will be detected in subgroup analyses therefore in addition to the smoking interaction analysis we intend to perform a diabetes-stratified analysis.

8. Discussion

8.1 Main outcomes of this thesis

In this thesis algorithms were developed to identify individuals with CAD, IS, and LEAD from EMR linked to the GoDARTS cohort. These algorithms were validated using data available from the GoDARTS cohort and known genetic associations in *CDKN2BA* and *LPA* regions were replicated in derived cardiovascular phenotypes. When genetic risk scores were associated with the cardiovascular phenotypes distinct association patterns with individual phenotypes were observed indicating that the derived cardiovascular phenotypes are distinct from each other despite an overlap in individuals. The genetic score for triglycerides showed an inverse association compared to the association serum triglyceride measures.

The CAD phenotype was used in the largest meta-analysis of CAD to date and identified 15 new loci for CAD and 104 loci that passed an FDR threshold of 5%. An CAD genetic risk score did not improve the prediction of CAD events above serum cholesterol levels in GoDARTS. In a large meta-analysis of CAD in T2D individuals only two independent signals in *ADAMTS7* which had previously been associated with CAD were detected at genome wide significance. There was significant allelic heterogeneity for the risk alleles in 6 of the known CAD loci indicating that there may be differences in effect between T2D and non-diabetic populations. We also detected rs944801 in the 9p21 region was associated with T2D and CAD in T2D, and had a nominally insignificant interaction with T2D. The risk of CAD was increased in individuals who had co-morbid type 2 diabetes.

In a Mendelian Randomisation study of CAD and vitamin D in GoDARTS, low vitamin D levels were associated with CAD and a genotype score for decreasing vitamin D levels was also found to be associated with CAD. These associations indicated that vitamin D was likely to be a risk factor for CAD. A meta-analysis of LEAD detected known signals in *CHRNA3* and the 9p21 region. The smoking interaction analysis detected a new locus for LEAD, *ADMATS17*. SNPs in this locus have different allelic effects in smokers vs. non-smokers.

8.2 The value and future applications of biobanks

Biobanks such as GoDARTS provide a platform for the study of many phenotypes for a relatively low cost compared to cross sectional studies. This is due to the comprehensive and rich data available for each study participant. In this thesis algorithms were developed to identify individuals with CAD, IS, and LEAD from EMR linked to the GoDARTS cohort. Similarly other phenotypes can be defined from the same group and can be investigated using the common genetic data available.

Other studies like the UK Biobank[356] and the Kaiser permanente/Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort[357] have a similar design to GoDARTS where DNA has been collected and linked to phenotype data. The UK BioBank has recruited 500,000, collected biological specimens and phenotype data for all its participants[356]. Currently the GERA cohort consists of 100,000 individuals with plans to genotype a further 400,000 individuals on high density SNP platforms. This cohort is linked to EMR data that allows for a wide range of phenotypes to be identified in a large population.

These studies are extremely valuable for identifying factors that are predictive of future disease due to their size and longitudinal data. Cross sectional studies are limited in this aspect due to their design as they essentially observe one slice of patient time and only observe living individuals[358]. These studies have been very useful in identifying loci that are associated with CAD but the predictive value of these variants is lacking compared to conventional risk factors (Chapter 4). Biobanks monitor individuals prospectively and collect additional phenotype data regularly over an individual's lifetime. Prospectively modelling changing cholesterol measures, blood pressure, weight, smoking status and prescribed medications is useful for designing predictive models for cardiovascular outcomes and for identifying genetic variants that predict future disease. Biobanks store a range of biological specimens from their study participants and these can be used to elucidate roles of SNPs in disease process and to identify biomarkers. Projects within SUMMIT are combining genetic data with proteomic and metabolomics data to identify new biomarkers for cardiovascular disease in diabetics. The SUMMIT study relies heavily on biobanks for suitable samples to analyse.

8.3 Explaining the missing heritability in CAD

Loci identified in large genetic studies explain 25% of the total heritability of CAD which means that 75% of the genetic components that the heritability have not been identified. Up until now the large-scale genetic association studies for CAD have focussed on identifying common variants that explain the additive variance of CAD while investigations of the non-additive variance such as epigenetic modifications, epistasis or gene by environment interactions have not been attempted on a large scale. The contribution of rare variants to the missing heritability has also not been quantified in large-scale experiments.

8.3.1 Identification of rare variants that explain the additive variance

The large consortia have proposed a series of 1000G imputations to identify rare variants associated with different traits. The reference panel for imputation has been assembled from the 1000 genomes project (http://www.1000genomes.org/) which includes whole genome sequencing of individuals from different Hap Map ethnic groups. This reference panel includes 2000 haplotypes for 46 Million polymorphisms for variants, insertions and deletions and sequence polymorphisms. This reference panel will be used to impute polymorphisms into existing genotype data that may explain some of the missing heritability for CAD. Other efforts to account for the missing heritability in common disease include large exome array experiments.

Similar to the CardioMetabochip, the exome chip is a custom array which includes non-synonymous, splice, stop altering, tags for previous GWAS hits, ancestry informative markers for African America, European and Native American ancestry, synonymous SNPs, mitochondrial SNPs, Y chromosome SNPs and HLA tags (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Currently large populations are being typed on these arrays for a series of exome chip analyses that will be performed. The success of analyses to identify rare variants and variants that have modest allelic effects is greatly dependent upon large populations to analyse and appropriate statistical analyses. The first publication of exome array analyses has detected rare variants that influence glucose homeostasis[380]. The exome analysis for CAD is currently underway and should provide interesting results.

Other efforts to identify rare variants associated with disease traits are underway. Whole genome sequencing projects like UK10K are well underway with the express purpose of understanding the genetic architecture of rare variants that determine different disease traits in 10 000 individuals. Other projects like GoT2D will also whole genome sequence 3000 individuals and 4000 individuals for exomes to identify rare variants associated with T2D. Sequencing projects are useful in identifying additional variants that are not in linkage with variants typed on large SNP genotyping arrays or that are rare or private mutations that contribute toward a disease.

8.3.2 Explaining the non-additive variance of common diseases

Limited analysis has been conducted to identify variants that interact with environmental factors to modify CAD risk. Stratified analyses performed in chapter 4 successfully identified rs16986953, a SNP that interacted with age to increase CAD risk in young CAD cases. Stratified analyses by risk factor have not been performed and we plan to perform a stratified analysis that includes non-diabetic individuals with CAD to build on work conducted in chapter 5. We will perform stratified analyses to identify loci that interact with diabetic status to modify CAD risk. GoDARTS is also taking part in a large smoking interaction analysis to identify variants that interact with smoking status to modify CAD risk.

Other avenues to quantify the non-additive variance in CAD include experiments that look for changes in the cells that are not due to changes in the DNA sequence or at how genes interact with other genes to produce an effect on CAD. Epigenetics is the study of changes in gene expression or cellular phenotypes that are not due to changes in DNA sequence. Epigenetics includes the study of methylation, histone modification and micro RNA alterations. These mechanisms help the cells to rapidly respond to environmental changes. Epigenetic alterations have been observed for CAD in response to nutrition, smoking, pollution, stress and circadian rhythm. These modification may explain some of the missing heritability and highlight new targets for therapy and disease prevention[381]. Gene by gene interactions are also important in explaining the missing heritability in CAD. Polymorphisms from individual genes may not be individually associated with CAD but may interact with polymorphisms in other genes to have an effect on CAD outcome such as polymorphisms in the renin-angiotensin system[382].

8.4 Analyses for IS and LEAD

In contrast with other phenotypes there are far fewer loci that have been identified for IS and LEAD. This may be due to smaller populations with genome wide array data for IS and LEAD. The stroke phenotype is also heterogeneous so the ISGC has moved its focus away from looking at the global IS phenotype to looking at TOAST stroke subtypes. Variants have been reported for large artery stroke and cardio-embolic stroke[143, 145] and larger studies focused on more distinct phenotypes may be more fruitful than previous analyses. The only large-scale meta-analysis published for LEAD lacked power to detect SNPs with small allelic effects[158] and much larger populations are required to identify additional variants. The meta-analysis of LEAD will be expanded to include more cohorts; we will also perform a diabetes-stratified analysis and will expand the smoking stratified analysis to include more samples.

8.5 Challenges for predicting CAD using genetic variants

The large CardioMetabochip replication of content contributed by large consortia is the first time that large populations have been typed on the same chip. As the chip was designed to replicate putative loci it is enriched with signals that are likely to be associated with targeted phenotypes. The cost of genotyping individuals on the CardioMetabochip allowed for more individuals to be genotyped increasing the power of analyses to detect SNPs with small allelic effects for individual phenotypes and to detect SNPs that have effects on multiple phenotypes. This provides a good resource for across phenotype comparisons and it is not surprising that many of the CAD associated SNPs were also associated with related risk factors: lipids, glycaemic traits, blood pressure and anthropometric measures[285]. Although the majority of variants that explain the total heritability of CAD have not been identified a large number of SNPs associated with CAD and related risk factors have been identified. How can we use the current body of known SNPs to predict CVD?

The prediction of CAD may be greatly improved by including SNPs associated with common CAD risk factors in addition to those identified as SNPs solely associated with CAD outcome. Not all SNPs will be useful instruments for predicting CAD outcomes so there needs to be a process to identify key SNPs that are useful as disease predictors. One of the major challenges in identifying a set of predictive SNPs is that some SNPs have pleiotropic effects where the effect of SNPs on the intermediate phenotypes is

not always consistent with increasing CAD risk. Mendelian randomisation studies can be useful in identifying SNPs that affect the intermediate phenotypes in a manner that is consistent with the allelic increase in CAD risk. Creating a composite score of CAD associated SNPs and those that are useful predictors of CAD risk factors and CAD outcomes may provide a sharper instrument to predict CAD outcome.

8.5.1 Investigating SNP pleiotropy

In chapter 3 we combined known variants for cardiovascular risk factors into genetic scores that were with CAD, IS and LEAD. Despite establishing that the genetic scores predicted individual risk factors, the effect of the genetic score for triglycerides was in opposition to the effect of serum triglycerides on cardiovascular outcomes. These conflicting associations are likely due to pleiotropic effects of SNPs that make up the triglyceride score. SNP pleiotropy is observed when SNPs occur in genes that have effects on multiple phenotypes.

It is important to consider that SNPs are associated with serum measures of triglycerides that are determined by a number of biological processes. SNPs that reduce serum triglyceride levels may do so by diverting triglycerides into surrounding tissues like the liver or the heart thus inducing lipotoxicity in vital organs of the body. Pleiotropy has been observed for rs1260326 (rs780094) that was included in the genetic score for triglycerides. The minor T-allele of rs780094 in GCKR raises triglyceride levels while simultaneously improving insulin sensitivity and raising HDL levels in individuals with metabolic syndrome[267, 279-281, 383]. Thus a genetic increase in triglycerides does not always correspond to an increase in disease risk and is not useful for disease prediction.

Identifying pleiotropic SNPs and investigating the mechanism behind the pleiotropic effects improves our knowledge of the genetic architecture of complex traits. It also informs our understanding of the underlying biology and phenotypic consequences of SNPs identified in large genetic studies. If we want to use variants identified in large genetic studies to predict disease we need to fully understand how that SNP affects the trait directly and how it affects intermediate phenotypes. Mendelian randomisation studies are commonly used to infer the causal relationship between an

outcome and an exposure. When relationships are well established it may be used to detect pleiotropic genetic effects in associated SNPs.

8.6 Mendelian randomisation studies

8.6.1 Inferring causal relationships

Mendelian randomisation has been applied successfully to confirming a causal relationship between interleukin-6[168] and CAD, and disproving a relationship between C-reactive protein[334] and CAD. In chapter 6 we applied a Mendelian randomisation design to determining the causal relationship between vitamin D levels and CAD. In GoDARTS we found a causal relationship between low vitamin D levels and CAD that has not been shown before. We have requested replication of these findings in other cohorts and have found interesting patterns in the data where the vitamin D lowering alleles correspond to an increase in CAD while in other cohorts closer to the equator the effects of the vitamin D lowering alleles tend to the null.

Due to the source of our vitamin D data we find that the mean vitamin D levels are much lower than those published by other studies. This may be due to the reduced daylight hours and sun exposure in Scotland and vitamin D measurements were ordered for medical investigations. When we compared our mean measured vitamin D per genotype with those published we found the same reduction in vitamin D per additional copy of the lowering allele. This would indicate that the SNPs have the same effect in our population but the basal vitamin D levels are naturally lower due to the environment. Based on these results we hypothesised that in populations that have naturally lower vitamin D levels due to increasing distance from the equator the effect of vitamin D reduction on cardiovascular outcome would be more pronounced.

In populations that are exposed to more sunlight and have healthy levels of vitamin D (70-250nmol/L)[198] (Figure 1-4) a reduction of 16nmol/L of vitamin D (Table 6-3), in individuals who are homozygous for all four of the vitamin D reducing alleles, is unlikely to result in adverse health outcomes. In populations that are have naturally lower vitamin D levels of less than 70nmol/L, which we observe for populations at greater than 50° from the equator (Figure 1-4), a 16nmol/L decrease in vitamin D due to genotype will cause vitamin D deficiency or insufficiency. It would be in these

populations that you would expect to see the greatest effect of vitamin D lowering SNPs on health outcomes.

These findings would provide a case for considering not just the effect of SNPs on intermediate phenotypes but also the context in which they are acting. These could include environmental or treatment factors that influence the intermediate phenotypes. Mendelian randomisation can be used to identify SNPs that affect an outcome through an intermediate phenotype. This approach was adopted by Freathy *et al.,* where rs9939609 in the *FTO* gene was tested for association with traits affected by BMI and it was shown that these relationships are consistent with the effect of *FTO* variants on BMI[341].

8.6.2 Identifying variants for a global CAD score from Mendelian randomisation studies

Similarly this method could be used to identify SNPs that predict CVD outcomes through their effects on lipid traits or other cardiovascular risk factors with the purpose of building a more global genetic prediction model for CAD. We could use a series of reductive Mendelian randomisation studies in large populations to identify the most suitable SNPs to be included in a predictive score for CAD. With the ultimate aim of predicting a lifetime risk of CAD based on carrying certain alleles.

In this thesis we have shown the known SNPs for CAD did not predict CAD better than serum lipid measures (Chapter 3). This would indicate that our current knowledge of CAD genetics cannot be used in a clinical setting. In order to use our knowledge to improve patient outcomes further studies need to be conducted to identify novel variants associated with CAD and follow up studies on the predictive ability for CAD outcomes of these variants need to be conducted.

9. Conclusions

Experiments, which identify the biological role and processes that these SNPs represent need to be conducted. A series of Mendelian randomisation studies need to be conducted on SNPs associated with cardiovascular risk factors to identify SNPs that are useful for predicting future CVD events. This thesis has highlighted how SNP x environment interactions need to be considered in Mendelian randomisation studies.

Much larger studies with focused analyses are required to detect both common and rare variants associated with CAD, IS and LEAD.

Publications

Published

1. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, Konig IR, Cazier JB, Johansson A, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikainen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, **Van Zuydam N**, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney AS, Mokhtari NE, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Muller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schafer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, Schoot CE, Wagner PJ, Wells GA, Wild PS, Yang TP, Amouyel P, Arveiler D, Basart H, Boehnke M, Boerwinkle E, Brambilla P, Cambien F, Cupples AL, de Faire U, Dehghan A, Diemert P, Epstein SE, Evans A, Ferrario MM, Ferrieres J, Gauguier D, Go AS, Goodall AH, Gudnason V, Hazen SL, Holm H, Iribarren C, Jang Y, Kahonen M, Kee F, Kim HS, Klopp N, Koenig W, Kratzer W, Kuulasmaa K, Laakso M, Laaksonen R, Lind L, Ouwehand WH, Parish S, Park JE, Pedersen NL, Peters A, Quertermous T, Rader DJ, Salomaa V, Schadt E, Shah SH, Sinisalo J, Stark K, Stefansson K, Tregouet DA, Virtamo J, Wallentin L, Wareham N, Zimmermann ME, Nieminen MS, Hengstenberg C, Sandhu MS, Pastinen T, Syvanen AC, Hovingh GK, Dedoussis G, Franks PW, Lehtimaki T, Metspalu A, Zalloua PA, Siegbahn A, Schreiber S, Ripatti S, Blankenberg SS, Perola M, Clarke R, Boehm BO, O'Donnell C, Reilly MP, Marz W, Collins R, Kathiresan S, Hamsten A, Kooner JS, Thorsteinsdottir U, Danesh J, Palmer CN, Roberts R, Watkins H, Schunkert H, Samani NJ. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2012;45:25-33

2. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, Fornage M, Ikram MA, Malik R, Bevan S, Thorsteinsdottir U, Nalls MA, Longstreth W, Wiggins KL, Yadav S, Parati EA, Destefano AL, Worrall BB, Kittner SJ, Khan MS, Reiner AP, Helgadottir A, Achterberg S, Fernandez-Cadenas I, Abboud S, Schmidt R, Walters M, Chen WM, Ringelstein EB, O'Donnell M, Ho WK, Pera J, Lemmens R, Norrving B, Higgins P, Benn M, Sale M, Kuhlenbaumer G, Doney AS, Vicente AM, Delavaran H, Algra A, Davies G, Oliveira SA, Palmer CN, Deary I, Schmidt H, Pandolfo M, Montaner J, Carty C, de Bakker PI, Kostulas K, Ferro JM, **van Zuydam NR**, Valdimarsson E, Nordestgaard BG, Lindgren A, Thijs V, Slowik A, Saleheen D, Pare G, Berger K, Thorleifsson G, Hofman A, Mosley TH, Mitchell BD, Furie K, Clarke R, Levi C, Seshadri S, Gschwendtner A, Boncoraglio GB, Sharma P, Bis JC, Gretarsdottir S, Psaty BM, Rothwell PM, Rosand J, Meschia JF, Stefansson K, Dichgans M, Markus HS. Genetic risk factors for ischaemic stroke and its subtypes (the metastroke collaboration): A meta-analysis of genome-wide association studies. *Lancet neurology*. 2012

3. Sobrin L, Green T, Sim X, Jensen RA, Tai ES, Tay WT, Wang JJ, Mitchell P, Sandholm N, Liu Y, Hietala K, Iyengar SK, Brooks M, Buraczynska M, **Van Zuydam N**, Smith AV, Gudnason V, Doney AS, Morris AD, Leese GP, Palmer CN, Swaroop A, Taylor HA, Jr., Wilson JG, Penman A, Chen CJ, Groop PH, Saw SM, Aung T, Klein BE, Rotter JI, Siscovick DS, Cotch MF, Klein R, Daly MJ, Wong TY. Candidate gene association study for diabetic retinopathy in persons with type 2 diabetes: The candidate gene association resource (care). *Invest Ophthalmol Vis Sci*. 2011

Manuscripts in progress

1. **Van Zuydam, N. R.**, Witham, M., Morris, A., Palmer, C.N.A and Doney, A., Single nucleotide polymorphisms influencing 25-hydroxyvitamin D levels have the expected effect on coronary artery disease events - a Mendelian randomization study in Go-DARTS. ATVB (under review)

2. **Van Zuydam, NR**, Donnelly LA, Hothersall EJ, Flynn R, Carr F, Tavendale R, Zhou K, Leese G, Morris AD, Doney ASF, Colhoun H, Palmer CNA. 2012. Using Electronic Medical Records to Investigate Cardiovascular Phenotype-Genotype Associations – A GoDARTS study. Circ. Card. Genetics (Submitted).

3. Fall, T.,Hägg, S., Mägi, R., Ploner, A., Fischer, K., Horikoshi, M., Sarin, A.P., Thorleifsson, G., Ladenvall, C., Kals, M., Kuningas, M., Draisma, H.M., Ried, J.S., **van Zuydam, N.R.** *et al.* Mendelian Randomization Analyses in ~200,000 Individuals Provide New Insights about the Role of Obesity in Cardiometabolic Traits. Under review

4. Leusink M, Onland-Moret NC, Asselbergs FW, Ding B, Kotti S, **van Zuydam NR**, Papp AC, Danchin N, Donnelly L, Morris AD, Chasman DI, Doevendans PAFM, Klungel OH, Ridker PM, van Gilst WH, Simon T, Nyberg F, Palmer CNA, Sadee W, Harst Pvd, Bakker PIWd, Boer Ad, Verstuyft C, Zee AHM-vd. Cholesteryl ester transfer protein (CETP) polymorphisms, statin use, and their impact on cholesterol levels and cardiovascular events. Submitted to JACC

5. Parry HM, Donnelly LA, **van Zuydam NR**, Doney AS, Elder DH, WTCCC2, Morris AD, Struthers AD, Palmer CN, Lang CC. Genetic variants predicting left ventricular hypertrophy in a diabetic population. Circ. Card. Genetics (Submitted).

6. Donnelly LA, **Van Zuydam N**, Zhou K, Tavendale R, Carr F, Zee A-HM-vd, Leusink M, de Boer A, Klungel OH, Doevendans PA, Morris AD, Pearson ER, Doney AS, Palmer CN. Robust association of the LPA locus with LDLC lowering response to statin treatment in a meta-analysis of 30,467 individuals from both randomised control trials and observational studies and association with coronary artery disease outcome during statin treatment.

References

1. Guo YD, Strugnell S, Back DW, Jones G. Transfected human liver cytochrome p-450 hydroxylates vitamin d analogs at different side-chain positions. *Proc Natl Acad Sci U S A*. 1993;90:8668-8672

2. Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, Jacobs EJ, Ascherio A, Helzlsouer K, Jacobs KB, Li Q, Weinstein SJ, Purdue M, Virtamo J, Horst R, Wheeler W, Chanock S, Hunter DJ, Hayes RB, Kraft P, Albanes D. Genome-wide association study of circulating vitamin d levels. *Hum Mol Genet*. 2010;19:2739-2745

3. Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW. Genetic evidence that the human cyp2r1 enzyme is a key vitamin d 25-hydroxylase. *Proc Natl Acad Sci U S A*. 2004;101:7711-7715

4. Waterham HR, Wanders RJ. Biochemical and genetic aspects of 7-dehydrocholesterol reductase and smith-lemli-opitz syndrome. *Biochim Biophys Acta*. 2000;1529:340-356

5. Norman AW. From vitamin d to hormone d: Fundamentals of the vitamin d endocrine system essential for good health. *Am J Clin Nutr*. 2008;88:491S-499S

6. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel DP, Streeten EA, Ohlsson C, Koller DL, Peltonen L, Cooper JD, O'Reilly PF, Houston DK, Glazer NL, Vandenput L, Peacock M, Shi J, Rivadeneira F, McCarthy MI, Anneli P, de Boer IH, Mangino M, Kato B, Smyth DJ, Booth SL, Jacques PF, Burke GL, Goodarzi M, Cheung CL, Wolf M, Rice K, Goltzman D, Hidiroglou N, Ladouceur M, Wareham NJ, Hocking LJ, Hart D, Arden NK, Cooper C, Malik S, Fraser WD, Hartikainen AL, Zhai G, Macdonald HM, Forouhi NG, Loos RJ, Reid DM, Hakim A, Dennison E, Liu Y, Power C, Stevens HE, Jaana L, Vasan RS, Soranzo N, Bojunga J, Psaty BM, Lorentzon M, Foroud T, Harris TB, Hofman A, Jansson JO, Cauley JA, Uitterlinden AG, Gibson Q, Jarvelin MR, Karasik D, Siscovick DS, Econs MJ, Kritchevsky SB, Florez JC, Todd JA, Dupuis J, Hypponen E, Spector TD. Common genetic determinants of vitamin d insufficiency: A genome-wide association study. *Lancet*. 2010;376:180-188

7. Ramagopalan SV, Heger A, Berlanga AJ, Maugeri NJ, Lincoln MR, Burrell A, Handunnetthi L, Handel AE, Disanto G, Orton SM, Watson CT, Morahan JM, Giovannoni G, Ponting CP, Ebers GC, Knight JC. A chip-seq defined genome-wide map of vitamin d receptor binding: Associations with disease and evolution. *Genome Res*. 2010;20:1352-1360

8. Davies MJ, Thomas AC. Plaque fissuring--the cause of acute myocardial infarction, sudden ischaemic death, and crescendo angina. *Br Heart J*. 1985;53:363-373

9. Aboyens V, Criqui MH. The epidemiology of peripheral aterial disease. In: Dieter RS, ed. *Peripheral arterial disease*. China: McGraw-Hill Companies; 2009:1.

10. Kannel WB, McGee DL. Diabetes and cardiovascular risk factors: The framingham study. *Circulation*. 1979;59:8-13

11. Haffner SM, Lehto S, Ronnemaa T, Pyorala K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med*. 1998;339:229-234

12. Stamler J, Vaccaro O, Neaton JD, Wentworth D. Diabetes, other risk factors, and 12-yr cardiovascular mortality for men screened in the multiple risk factor intervention trial. *Diabetes Care*. 1993;16:434-444

13. Gomes M, Giannella-Neto D, Faria M, Tambascia M, Fonseca R, Rea R, Macedo G, Modesto-Filho J, Schmid H, Bittencourt A, Cavalcanti S, Rassi N, Pedrosa H, Dib S. Estimating cardiovascular risk in patients with type 2 diabetes: A national multicenter study in brazil. *Diabetology & Metabolic Syndrome*. 2009;1:22

14. Bucala R, Makita Z, Vega G, Grundy S, Koschinsky T, Cerami A, Vlassara H. Modification of low-density-lipoprotein by advanced glycation end-products contributes to the dyslipidemia of diabetes and renal-insufficiency. *Proceedings of the National Academy of Sciences of the United States of America*. 1994;91:9441-9445

15. Bucala R. Lipoprotein modification by advanced glycosylation endproducts (ages): Role in atherosclerosis. *Trends in cardiovascular medicine*. 1997;7:39-47

16. Hayden M, Tyagi S. Intimal redox stress: Accelerated atherosclerosis in metabolic syndrome and type 2 diabetes mellitus. Atherosclеropathy. *Cardiovascular Diabetology*. 2002;1:3

17. Lusis A. Atherosclerosis. *NATURE*. 2000;407:233-241

18. Newby AC, Zaltsman AB. Fibrous cap formation or destruction--the critical importance of vascular smooth muscle cell proliferation, migration and matrix formation. *Cardiovasc Res*. 1999;41:345-360

19. Frid MG, Dempsey EC, Durmowicz AG, Stenmark KR. Smooth muscle cell heterogeneity in pulmonary and systemic vessels. Importance in vascular disease. *Arterioscler Thromb Vasc Biol*. 1997;17:1203-1209

20. Ross R, Glomset J, Harker L. Response to injury and atherogenesis. *The American journal of pathology*. 1977;86:675-684

21. Lusis AJ. Atherosclerosis. *NATURE*. 2000;407:233-241

22. Williams KJ, Tabas I. The response-to-retention hypothesis of early atherogenesis. *Arterioscler Thromb Vasc Biol*. 1995;15:551-561

23. Grossman LK, Harter C, Hasbrouck C. Testing mothers' knowledge of breastfeeding: Instrument development and implementation and correlation with infant feeding decision. *Journal of pediatric & perinatal nutrition*. 1990;2:43-63

24. Tedgui A, Mallat Z. [atherosclerotic plaque formation]. *La Revue du praticien*. 1999;49:2081-2086

25. Munzel T, Sinning C, Post F, Warnholtz A, Schulz E. Pathophysiology, diagnosis and prognostic implications of endothelial dysfunction. *Annals of medicine*. 2008;40:180-196

26. Ross R. The pathogenesis of atherosclerosis - a perspective for the 1990s. *Nature*. 1993;362:801-809

27. Masuda J, Ross R. Atherogenesis during low level hypercholesterolemia in the nonhuman primate. Ii. Fatty streak conversion to fibrous plaque. *Arteriosclerosis*. 1990;10:178-187

28. Hayden M, Tyagi S. "A" is for amylin and amyloid in type 2 diabetes mellitus. *JOP. J Pancreas (Online)*. 2001;2:124 - 139

29. Hayden M, Tyagi S. Islet redox stress: The manifold toxicities of insulin resistance, metabolic syndrome, and amylin derived islet amyloid in type 2 diabetes mellitus. *JOP. J Pancreas (Online)*. 2002;3:86 - 108

30. Bierman EL. George lyman duff memorial lecture. Atherogenesis in diabetes. *Arteriosclerosis and thrombosis : a journal of vascular biology / American Heart Association*. 1992;12:647-656

31. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, Pihur V, Vollenweider P, O'Reilly PF, Amin N, Bragg-Gresham JL, Teumer A, Glazer NL, Launer L, Zhao JH, Aulchenko Y, Heath S, Sober S, Parsa A, Luan J, Arora P, Dehghan A, Zhang F, Lucas G, Hicks AA, Jackson AU, Peden JF, Tanaka T, Wild SH, Rudan I, Igl W, Milaneschi Y, Parker AN, Fava C, Chambers JC, Fox ER, Kumari M, Go MJ, van der Harst P, Kao WH, Sjogren M, Vinay DG, Alexander M, Tabara Y, Shaw-Hawkins S, Whincup PH, Liu Y, Shi G, Kuusisto J, Tayo B, Seielstad M, Sim X, Nguyen KD, Lehtimaki T, Matullo G, Wu Y, Gaunt TR, Onland-Moret NC, Cooper MN, Platou CG, Org E, Hardy R, Dahgam S, Palmen J, Vitart V, Braund PS, Kuznetsova T, Uiterwaal CS, Adeyemo A, Palmas W, Campbell H, Ludwig B, Tomaszewski M, Tzoulaki I, Palmer ND, Aspelund T, Garcia M, Chang YP, O'Connell JR, Steinle NI, Grobbee DE, Arking DE, Kardia SL, Morrison AC, Hernandez D, Najjar S, McArdle WL, Hadley D, Brown MJ, Connell JM, Hingorani AD, Day IN, Lawlor DA, Beilby JP, Lawrence RW, Clarke R, Hopewell JC, Ongen H, Dreisbach AW, Li Y, Young JH, Bis JC, Kahonen M, Viikari J, Adair LS, Lee NR, Chen MH, Olden M, Pattaro C, Bolton JA, Kottgen A, Bergmann S, Mooser V, Chaturvedi N, Frayling TM, Islam M, Jafar TH, Erdmann J, Kulkarni SR, Bornstein SR, Grassler J, Groop L, Voight BF, Kettunen J, Howard P, Taylor A, Guarrera S, Ricceri F, Emilsson V, Plump A, Barroso I, Khaw KT, Weder AB, Hunt SC, Sun YV, Bergman RN, Collins FS, Bonnycastle LL, Scott LJ, Stringham HM, Peltonen L, Perola M, Vartiainen E, Brand SM, Staessen JA, Wang TJ, Burton PR, Artigas MS, Dong Y, Snieder H, Wang X, Zhu H, Lohman KK, Rudock ME, Heckbert SR, Smith NL, Wiggins KL, Doumatey A, Shriner D, Veldre G, Viigimaa M, Kinra S, Prabhakaran D, Tripathy V, Langefeld CD, Rosengren A, Thelle DS, Corsi AM, Singleton A, Forrester T, Hilton G, McKenzie CA, Salako T, Iwai N, Kita Y, Ogihara T, Ohkubo T, Okamura T, Ueshima H, Umemura S, Eyheramendy S, Meitinger T, Wichmann HE, Cho YS, Kim HL, Lee JY, Scott J, Sehmi JS, Zhang W, Hedblad B, Nilsson P, Smith GD, Wong A, Narisu N, Stancakova A, Raffel LJ, Yao J, Kathiresan S, O'Donnell CJ, Schwartz SM, Ikram MA, Longstreth WT, Jr., Mosley TH, Seshadri S, Shrine NR, Wain LV, Morken MA, Swift AJ, Laitinen J, Prokopenko I, Zitting P, Cooper JA, Humphries SE, Danesh J, Rasheed A, Goel A, Hamsten A, Watkins H, Bakker SJ, van Gilst WH, Janipalli CS, Mani KR, Yajnik CS, Hofman A, Mattace-Raso FU, Oostra BA, Demirkan A, Isaacs A, Rivadeneira F, Lakatta EG, Orru M, Scuteri A, Ala-Korpela M, Kangas AJ, Lyytikainen LP, Soininen P, Tukiainen T, Wurtz P, Ong RT, Dorr M, Kroemer HK, Volker U, Volzke H, Galan P,

Hercberg S, Lathrop M, Zelenika D, Deloukas P, Mangino M, Spector TD, Zhai G, Meschia JF, Nalls MA, Sharma P, Terzic J, Kumar MV, Denniff M, Zukowska-Szczechowska E, Wagenknecht LE, Fowkes FG, Charchar FJ, Schwarz PE, Hayward C, Guo X, Rotimi C, Bots ML, Brand E, Samani NJ, Polasek O, Talmud PJ, Nyberg F, Kuh D, Laan M, Hveem K, Palmer LJ, van der Schouw YT, Casas JP, Mohlke KL, Vineis P, Raitakari O, Ganesh SK, Wong TY, Tai ES, Cooper RS, Laakso M, Rao DC, Harris TB, Morris RW, Dominiczak AF, Kivimaki M, Marmot MG, Miki T, Saleheen D, Chandak GR, Coresh J, Navis G, Salomaa V, Han BG, Zhu X, Kooner JS, Melander O, Ridker PM, Bandinelli S, Gyllensten UB, Wright AF, Wilson JF, Ferrucci L, Farrall M, Tuomilehto J, Pramstaller PP, Elosua R, Soranzo N, Sijbrands EJ, Altshuler D, Loos RJ, Shuldiner AR, Gieger C, Meneton P, Uitterlinden AG, Wareham NJ, Gudnason V, Rotter JI, Rettig R, Uda M, Strachan DP, Witteman JC, Hartikainen AL, Beckmann JS, Boerwinkle E, Vasan RS, Boehnke M, Larson MG, Jarvelin MR, Psaty BM, Abecasis GR, Chakravarti A, Elliott P, van Duijn CM, Newton-Cheh C, Levy D, Caulfield MJ, Johnson T. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *NATURE*. 2011;478:103-109

32. CARDIoGRAM+C4D. A genome-wide association study in europeans and south asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43:339-344

33. Preuss M, Konig IR, Thompson JR, Erdmann J, Absher D, Assimes TL, Blankenberg S, Boerwinkle E, Chen L, Cupples LA, Hall AS, Halperin E, Hengstenberg C, Holm H, Laaksonen R, Li M, Marz W, McPherson R, Musunuru K, Nelson CP, Burnett MS, Epstein SE, O'Donnell CJ, Quertermous T, Rader DJ, Roberts R, Schillert A, Stefansson K, Stewart AF, Thorleifsson G, Voight BF, Wells GA, Ziegler A, Kathiresan S, Reilly MP, Samani NJ, Schunkert H. Design of the coronary artery disease genome-wide replication and meta-analysis (cardiogram) study: A genome-wide association meta-analysis involving more than 22 000 cases and 60 000 controls. *Circ Cardiovasc Genet*. 2010;3:475-483

34. Go AS, Iribarren C, Chandra M, Lathon PV, Fortmann SP, Quertermous T, Hlatky MA. Statin and beta-blocker therapy and the initial presentation of coronary heart disease. *Ann Intern Med*. 2006;144:229-238

35. Taylor-Piliae RE, Norton LC, Haskell WL, Mahbouda MH, Fair JM, Iribarren C, Hlatky MA, Go AS, Fortmann SP. Validation of a new brief physical activity survey among men and women aged 60-69 years. *Am J Epidemiol*. 2006;164:598-606

36. Iribarren C, Go AS, Husson G, Sidney S, Fair JM, Quertermous T, Hlatky MA, Fortmann SP. Metabolic syndrome and early-onset coronary artery disease: Is the whole greater than its parts? *J Am Coll Cardiol*. 2006;48:1800-1807

37. Assimes TL, Knowles JW, Basu A, Iribarren C, Southwick A, Tang H, Absher D, Li J, Fair JM, Rubin GD, Sidney S, Fortmann SP, Go AS, Hlatky MA, Myers RM, Risch N, Quertermous T. Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic advance study. *Hum Mol Genet*. 2008;17:2320-2328

38. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD,

Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443-453

39. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, Uitterlinden AG, Harris TB, Witteman JCM, Boerwinkle E, Consortium oBotC. Cohorts for heart and aging research in genomic epidemiology (charge) consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circulation: Cardiovascular Genetics*. 2009;2:73-80

40. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-1493

41. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet DA, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, El Mokhtari NE, Schafer A, Marz W, Renner W, Bugert P, Kluter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet*. 2009;41:280-282

42. Winkelmann BR, Marz W, Boehm BO, Zotz R, Hager J, Hellstern P, Senges J. Rationale and design of the luric study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics*. 2001;2:S1-73

43. Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, Mehta NN, Burnett MS, Devaney JM, Knouff CW, Thompson JR, Horne BD, Stewart AF, Assimes TL, Wild PS, Allayee H, Nitschke PL, Patel RS, Martinelli N, Girelli D, Quyyumi AA, Anderson JL, Erdmann J, Hall AS, Schunkert H, Quertermous T, Blankenberg S, Hazen SL, Roberts R, Kathiresan S, Samani NJ, Epstein SE, Rader DJ. Identification of adamts7 as a novel locus for coronary atherosclerosis and association of abo with myocardial infarction in the presence of coronary atherosclerosis: Two genome-wide association studies. *Lancet*. 2011;377:383-392

44. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, Berzuini C, Bernardinelli L, Peyvandi F, Tubaro M, Celli P, Ferrario M, Fetiveau R, Marziliano N, Casari G, Galli M, Ribichini F, Rossi M, Bernardi F, Zonzin P, Piazza A, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Daly MJ, Nemesh J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burtt N, Gabriel SB,

Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall A, Linsel-Nitschke P, Lieb W, Ziegler A, Konig I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Knouff CW, Waterworth DM, Walker MC, Mooser V, Epstein SE, Scheffold T, Berger K, Huge A, Martinelli N, Olivieri O, Corrocher R, McKeown P, Erdmann E, Konig IR, Holm H, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Do R, Xie C, Siscovick D. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009;41:334-341

45. Stewart AF, Dandona S, Chen L, Assogba O, Belanger M, Ewart G, LaRose R, Doelle H, Williams K, Wells GA, McPherson R, Roberts R. Kinesin family member 6 variant trp719arg does not associate with angiographically defined coronary artery disease in the ottawa heart genomics study. *J Am Coll Cardiol*. 2009;53:1471-1472

46. Lehrke M, Millington SC, Lefterova M, Cumaranatunge RG, Szapary P, Wilensky R, Rader DJ, Lazar MA, Reilly MP. Cxcl16 is a marker of inflammation, atherosclerosis, and acute coronary syndromes in humans. *J Am Coll Cardiol*. 2007;49:442-449

47. WTCCC2. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *NATURE*. 2007;447:661-678

48. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoep JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333-338

49. Consortium CADCDG. A genome-wide association study in europeans and south asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43:339-344

50. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, Bennett D, Silveira A, Malarstig A, Green FR, Lathrop M, Gigante B, Leander K, de Faire U, Seedorf U, Hamsten A, Collins R, Watkins H, Farrall M. Genetic variants associated with lp(a) lipoprotein level and coronary disease. *N Engl J Med*. 2009;361:2518-2528

51. Erdmann J, Groszhennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet D-A, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, Mokhtari NEE, Schafer A, Marz W, Renner W, Bugert P, Kluter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet*. 2009;41:280-282

52. Consortium IKC. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet*. 2011;7:e1002260

53. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet*. 2008;40:189-197

54. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, Konig IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI,

Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA, Jr., Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *NATURE*. 2010;466:707-713

55. Holzel M, Rohrmoser M, Schlee M, Grimm T, Harasim T, Malamoussi A, Gruber-Eber A, Kremmer E, Hiddemann W, Bornkamm GW, Eick D. Mammalian wdr12 is a novel member of the pes1-bop1 complex and is required for ribosome biogenesis and cell proliferation. *J Cell Biol*. 2005;170:367-378

56. Kimmelman AC, Nunez Rodriguez N, Chan AM. R-ras3/m-ras induces neuronal differentiation of pc12 cells through cell-type-specific activation of the mitogen-activated protein kinase cascade. *Molecular and cellular biology*. 2002;22:5946-5961

57. Jarray R, Allain B, Borriello L, Biard D, Loukaci A, Larghero J, Hadj-Slimane R, Garbay C, Lepelletier Y, Raynaud F. Depletion of the novel protein phactr-1 from human endothelial cells abolishes tube formation and induces cell death receptor apoptosis. *Biochimie*. 2011;93:1668-1675

58. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, Lauretani F, Murray A, Gibbs JR, Paolisso G, Rafiq S, Simon-Sanchez J, Lango H, Scholz S, Weedon MN, Arepalli S, Rice N, Washecka N, Hurst A, Britton A, Henley W, van de Leemput J, Li R, Newman AB, Tranah G, Harris T, Panicker V, Dayan C, Bennett A, McCarthy MI, Ruokonen A, Jarvelin MR, Guralnik J, Bandinelli S, Frayling TM, Singleton A, Ferrucci L. A genome-wide association study identifies protein quantitative trait loci (pqtls). *PLoS Genet*. 2008;4:e1000072

59. Wrensch M, Jenkins RB, Chang JS, Yeh RF, Xiao Y, Decker PA, Ballman KV, Berger M, Buckner JC, Chang S, Giannini C, Halder C, Kollmeyer TM, Kosel ML, LaChance DH, McCoy L, O'Neill BP, Patoka J, Pico AR, Prados M, Quesenberry C, Rice T, Rynearson AL, Smirnov I, Tihan T, Wiemels J, Yang P, Wiencke JK. Variants in the cdkn2b and rtel1 regions are associated with high-grade glioma susceptibility. *Nat Genet*. 2009;41:905-908

60. Wehler TC, Graf C, Altherr K, Zimmermann T, Brenner W, Thuroff JW, Biesterfeld S, Gockel I, Theobald M, Galle PR, Schimanski CC. Sdf1beta expression in renal cell carcinoma correlates with grading and infiltration by cd8+ t-cells. *Anticancer research*. 2011;31:2797-2803

61. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen SC, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfort J, Gupta P, Huether R, Kriwacki RW, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, Cheng C, Roberts S, Barbato MI, Campana D, Coustan-Smith E, Shurtleff SA, Raimondi SC, Kleppe M, Cools J, Shimano KA, Hermiston ML, Doulatov S, Eppert K, Laurenti E, Notta F, Dick JE, Basso G, Hunger SP, Loh ML, Devidas M, Wood B, Winter S, Dunsmore KP, Fulton RS, Fulton LL, Hong X, Harris CC, Dooling DJ, Ochoa K, Johnson KJ, Obenauer JC, Evans WE, Pui CH, Naeve CW, Ley TJ, Mardis ER, Wilson RK,

Downing JR, Mullighan CG. The genetic basis of early t-cell precursor acute lymphoblastic leukaemia. *NATURE*. 2012;481:157-163

62. Plagnol V, Howson JM, Smyth DJ, Walker N, Hafler JP, Wallace C, Stevens H, Jackson L, Simmonds MJ, Bingley PJ, Gough SC, Todd JA. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet*. 2011;7:e1002216

63. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, Papadakis K, Voight BF, Scott LJ, Zhang F, Farrall M, Tanaka T, Wallace C, Chambers JC, Khaw KT, Nilsson P, van der Harst P, Polidoro S, Grobbee DE, Onland-Moret NC, Bots ML, Wain LV, Elliott KS, Teumer A, Luan J, Lucas G, Kuusisto J, Burton PR, Hadley D, McArdle WL, Brown M, Dominiczak A, Newhouse SJ, Samani NJ, Webster J, Zeggini E, Beckmann JS, Bergmann S, Lim N, Song K, Vollenweider P, Waeber G, Waterworth DM, Yuan X, Groop L, Orho-Melander M, Allione A, Di Gregorio A, Guarrera S, Panico S, Ricceri F, Romanazzi V, Sacerdote C, Vineis P, Barroso I, Sandhu MS, Luben RN, Crawford GJ, Jousilahti P, Perola M, Boehnke M, Bonnycastle LL, Collins FS, Jackson AU, Mohlke KL, Stringham HM, Valle TT, Willer CJ, Bergman RN, Morken MA, Doring A, Gieger C, Illig T, Meitinger T, Org E, Pfeufer A, Wichmann HE, Kathiresan S, Marrugat J, O'Donnell CJ, Schwartz SM, Siscovick DS, Subirana I, Freimer NB, Hartikainen AL, McCarthy MI, O'Reilly PF, Peltonen L, Pouta A, de Jong PE, Snieder H, van Gilst WH, Clarke R, Goel A, Hamsten A, Peden JF, Seedorf U, Syvanen AC, Tognoni G, Lakatta EG, Sanna S, Scheet P, Schlessinger D, Scuteri A, Dorr M, Ernst F, Felix SB, Homuth G, Lorbeer R, Reffelmann T, Rettig R, Volker U, Galan P, Gut IG, Hercberg S, Lathrop GM, Zelenika D, Deloukas P, Soranzo N, Williams FM, Zhai G, Salomaa V, Laakso M, Elosua R, Forouhi NG, Volzke H, Uiterwaal CS, van der Schouw YT, Numans ME, Matullo G, Navis G, Berglund G, Bingham SA, Kooner JS, Connell JM, Bandinelli S, Ferrucci L, Watkins H, Spector TD, Tuomilehto J, Altshuler D, Strachan DP, Laan M, Meneton P, Wareham NJ, Uda M, Jarvelin MR, Mooser V, Melander O, Loos RJ, Elliott P, Abecasis GR, Caulfield M, Munroe PB. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet*. 2009;41:666-676

64. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, Hwang JY, Oh JH, Kim DJ, Kim NH, Kim S, Hong EJ, Kim JH, Min H, Kim Y, Zhang R, Jia W, Okada Y, Takahashi A, Kubo M, Tanaka T, Kamatani N, Matsuda K, Park T, Oh B, Kimm K, Kang D, Shin C, Cho NH, Kim HL, Han BG, Cho YS. Large-scale genome-wide association studies in east asians identify new genetic loci influencing metabolic traits. *Nat Genet*. 2011;43:990-995

65. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, Albrecht E, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliovaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y, Loehr LR, Lohman K, Loos RJ, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin AC, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T,

Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietilainen KH, Pin I, Pola Sbreve Ek O, Pouta A, Psaty BM, Hartikainen AL, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin SY, Spector TD, Surakka I, Vitart V, Volzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N, Kaprio J, Wilson JF, Hayward C, Kahonen M, Heinrich J, Musk AW, Jarvis DL, Glaser S, Jarvelin MR, Ch Stricker BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43:1082-1090

66. Humtsoe JO, Liu M, Malik AB, Wary KK. Lipid phosphate phosphatase 3 stabilization of beta-catenin induces endothelial cell migration and formation of branching point structures. *Molecular and cellular biology*. 2010;30:1593-1606

67. Richards KL, Zhang B, Sun M, Dong W, Churchill J, Bachinski LL, Wilson CD, Baggerly KA, Yin G, Hayes DN, Wistuba, II, Krahe R. Methylation of the candidate biomarker tcf21 is very frequent across a spectrum of early-stage nonsmall cell lung cancers. *Cancer*. 2011;117:606-617

68. Paterson AD, Lopes-Virella MF, Waggott D, Boright AP, Hosseini SM, Carter RE, Shen E, Mirea L, Bharaj B, Sun L, Bull SB. Genome-wide association identifies the abo blood group as a major locus associated with serum levels of soluble e-selectin. *Arterioscler Thromb Vasc Biol*. 2009;29:1958-1967

69. Barbalic M, Dupuis J, Dehghan A, Bis JC, Hoogeveen RC, Schnabel RB, Nambi V, Bretler M, Smith NL, Peters A, Lu C, Tracy RP, Aleksic N, Heeriga J, Keaney JF, Jr., Rice K, Lip GY, Vasan RS, Glazer NL, Larson MG, Uitterlinden AG, Yamamoto J, Durda P, Haritunians T, Psaty BM, Boerwinkle E, Hofman A, Koenig W, Jenny NS, Witteman JC, Ballantyne C, Benjamin EJ. Large-scale genomic studies reveal central role of abo in sp-selectin and sicam-1 levels. *Hum Mol Genet*. 2010;19:1863-1872

70. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N. Genome-wide association study of hematological and biochemical traits in a japanese population. *Nat Genet*. 2010;42:210-215

71. Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, Sim X, Tay WT, Chen CH, Zhang Y, Yamamoto K, Katsuya T, Yokota M, Kim YJ, Ong RT, Nabika T, Gu D, Chang LC, Kokubo Y, Huang W, Ohnaka K, Yamori Y, Nakashima E, Jaquish CE, Lee JY, Seielstad M, Isono M, Hixson JE, Chen YT, Miki T, Zhou X, Sugiyama T, Jeon JP, Liu JJ, Takayanagi R, Kim SS, Aung T, Sung YJ, Zhang X, Wong TY, Han BG, Kobayashi S, Ogihara T, Zhu D, Iwai N, Wu JY, Teo YY, Tai ES, Cho YS, He J. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east asians. *Nat Genet*. 2011;43:531-538

72. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43:969-976

73. Kraja AT, Vaidya D, Pankow JS, Goodarzi MO, Assimes TL, Kullo IJ, Sovio U, Mathias RA, Sun YV, Franceschini N, Absher D, Li G, Zhang Q, Feitosa MF, Glazer NL, Haritunians T,

Hartikainen AL, Knowles JW, North KE, Iribarren C, Kral B, Yanek L, O'Reilly PF, McCarthy MI, Jaquish C, Couper DJ, Chakravarti A, Psaty BM, Becker LC, Province MA, Boerwinkle E, Quertermous T, Palotie L, Jarvelin MR, Becker DM, Kardia SL, Rotter JI, Chen YD, Borecki IB. A bivariate genome-wide approach to metabolic syndrome: Stampeed consortium. *Diabetes*. 2011;60:1329-1339

74. Volonghi I, Pezzini A, Del Zotto E, Giossi A, Costa P, Ferrari D, Padovani A. Role of col4a1 in basement-membrane integrity and cerebral small-vessel disease. The col4a1 stroke syndrome. *Current medicinal chemistry*. 2010;17:1317-1324

75. Liu CJ, Kong W, Ilalov K, Yu S, Xu K, Prazak L, Fajardo M, Sehgal B, Di Cesare PE. Adamts-7: A metalloproteinase that directly binds to and degrades cartilage oligomeric matrix protein. *FASEB J*. 2006;20:988-990

76. Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, Watzinger N, Larson MG, Smith NL, Dehghan A, Grosshennig A, Schillert A, Teumer A, Schmidt R, Kathiresan S, Lumley T, Aulchenko YS, Konig IR, Zeller T, Homuth G, Struchalin M, Aragam J, Bis JC, Rivadeneira F, Erdmann J, Schnabel RB, Dorr M, Zweiker R, Lind L, Rodeheffer RJ, Greiser KH, Levy D, Haritunians T, Deckers JW, Stritzke J, Lackner KJ, Volker U, Ingelsson E, Kullo I, Haerting J, O'Donnell CJ, Heckbert SR, Stricker BH, Ziegler A, Reffelmann T, Redfield MM, Werdan K, Mitchell GF, Rice K, Arnett DK, Hofman A, Gottdiener JS, Uitterlinden AG, Meitinger T, Blettner M, Friedrich N, Wang TJ, Psaty BM, van Duijn CM, Wichmann HE, Munzel TF, Kroemer HK, Benjamin EJ, Rotter JI, Witteman JC, Schunkert H, Schmidt H, Volzke H, Blankenberg S. Genetic variants associated with cardiac structure and function: A meta-analysis and replication of genome-wide association data. *JAMA*. 2009;302:168-178

77. Tsai FJ, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, Wang TY, Chen RH, Shiu CF, Liu YM, Chang CC, Chen P, Chen CH, Fann CS, Chen YT, Wu JY. A genome-wide association study identifies susceptibility variants for type 2 diabetes in han chinese. *PLoS Genet*. 2010;6:e1000847

78. Wagsater D, Zhu C, Bjorck HM, Eriksson P. Effects of pdgf-c and pdgf-d on monocyte migration and mmp-2 and mmp-9 expression. *Atherosclerosis*. 2009;202:415-423

79. Akashi M, Higashi T, Masuda S, Komori T, Furuse M. A coronary artery disease-associated gene product, jcad/kiaa1462, is a novel component of endothelial cell-cell junctions. *Biochem Biophys Res Commun*. 2011;413:224-229

80. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. Loci related to metabolic-syndrome pathways including lepr,hnf1a, il6r, and gckr associate with plasma c-reactive protein: The women's genome health study. *Am J Hum Genet*. 2008;82:1185-1192

81. Bots ML, Hoes AW, Koudstaal PJ, Hofman A, Grobbee DE. Common carotid intima-media thickness and risk of stroke and myocardial infarction: The rotterdam study. *Circulation*. 1997;96:1432-1437

82. Rappold E. [how do we experience our nursing education--what do we wish for?]. *Osterreichische Krankenpflegezeitschrift*. 1991;44 Suppl:90-92

83. Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B, Heuschmann PU. Epidemiology of ischemic stroke subtypes according to toast criteria: Incidence, recurrence, and long-term survival in ischemic stroke subtypes: A population-based study. *Stroke*. 2001;32:2735-2740

84. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, Meitinger T, Wichmann E, Mitchell BD, Furie K, Slowik A, Rich SS, Syme PD, MacLeod MJ, Meschia JF, Rosand J, Kittner SJ, Markus HS, Muller-Myhsok B, Dichgans M. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Annals of neurology*. 2009;65:531-539

85. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, Kostulas K, Kuhlenbaumer G, Bevan S, Jonsdottir T, Bjarnason H, Saemundsdottir J, Palsson S, Arnar DO, Holm H, Thorgeirsson G, Valdimarsson EM, Sveinbjornsdottir S, Gieger C, Berger K, Wichmann HE, Hillert J, Markus H, Gulcher JR, Ringelstein EB, Kong A, Dichgans M, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Annals of neurology*. 2008;64:402-409

86. Taira N, Ura S. Sudden death in calves associated with strongyloides papillosus infection. *Veterinary parasitology*. 1991;39:313-319

87. Kirchhof P, Kahr PC, Kaese S, Piccini I, Vokshi I, Scheld HH, Rotering H, Fortmueller L, Laakmann S, Verheule S, Schotten U, Fabritz L, Brown NA. Pitx2c is expressed in the adult left atrium, and reducing pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression. *Circ Cardiovasc Genet*. 2011;4:123-133

88. Benjamin EJ, Rice KM, Arking DE, Pfeufer A, van Noord C, Smith AV, Schnabel RB, Bis JC, Boerwinkle E, Sinner MF, Dehghan A, Lubitz SA, D'Agostino RB, Sr., Lumley T, Ehret GB, Heeringa J, Aspelund T, Newton-Cheh C, Larson MG, Marciante KD, Soliman EZ, Rivadeneira F, Wang TJ, Eiriksdottir G, Levy D, Psaty BM, Li M, Chamberlain AM, Hofman A, Vasan RS, Harris TB, Rotter JI, Kao WH, Agarwal SK, Stricker BH, Wang K, Launer LJ, Smith NL, Chakravarti A, Uitterlinden AG, Wolf PA, Sotoodehnia N, Kottgen A, van Duijn CM, Meitinger T, Mueller M, Perz S, Steinbeck G, Wichmann HE, Lunetta KL, Heckbert SR, Gudnason V, Alonso A, Kaab S, Ellinor PT, Witteman JC. Variants in zfhx3 are associated with atrial fibrillation in individuals of european ancestry. *Nat Genet*. 2009;41:879-881

89. Zhang K, Bai P, Shi S, Zhou B, Wang Y, Song Y, Rao L, Zhang L. The g894t polymorphism on endothelial nitric oxide synthase gene is associated with increased coronary heart disease among asia population: Evidence from a meta analysis. *Thrombosis research*. 2012;130:192-197

90. Bis JC, Kavousi M, Franceschini N, Isaacs A, Abecasis GR, Schminke U, Post WS, Smith AV, Cupples LA, Markus HS, Schmidt R, Huffman JE, Lehtimaki T, Baumert J, Munzel T, Heckbert SR, Dehghan A, North K, Oostra B, Bevan S, Stoegerer EM, Hayward C, Raitakari O, Meisinger C, Schillert A, Sanna S, Volzke H, Cheng YC, Thorsson B, Fox CS, Rice K, Rivadeneira F, Nambi V, Halperin E, Petrovic KE, Peltonen L, Wichmann HE, Schnabel RB, Dorr M, Parsa A, Aspelund T, Demissie S, Kathiresan S, Reilly MP, Taylor K, Uitterlinden A, Couper DJ, Sitzer M, Kahonen M, Illig T, Wild PS, Orru M, Ludemann J, Shuldiner AR, Eiriksdottir G, White CC,

Rotter JI, Hofman A, Seissler J, Zeller T, Usala G, Ernst F, Launer LJ, D'Agostino RB, Sr., O'Leary DH, Ballantyne C, Thiery J, Ziegler A, Lakatta EG, Chilukoti RK, Harris TB, Wolf PA, Psaty BM, Polak JF, Li X, Rathmann W, Uda M, Boerwinkle E, Klopp N, Schmidt H, Wilson JF, Viikari J, Koenig W, Blankenberg S, Newman AB, Witteman J, Heiss G, Duijn C, Scuteri A, Homuth G, Mitchell BD, Gudnason V, O'Donnell CJ. Meta-analysis of genome-wide association studies from the charge consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet*. 2011;43:940-947

91. Ding H, Xu Y, Bao X, Wang X, Cui G, Wang W, Hui R, Wang DW. Confirmation of genomewide association signals in chinese han population reveals risk loci for ischemic stroke. *Stroke*. 2010;41:177-180

92. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: The framingham study. *Stroke*. 1991;22:983-988

93. Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, Soranzo N, Whittaker P, Ranganath V, Kumanduri V, McLaren W, Holm L, Lindh J, Rane A, Wadelius M, Deloukas P. A genome-wide association study confirms vkorc1, cyp2c9, and cyp4f2 as principal genetic determinants of warfarin dose. *PLoS Genet*. 2009;5:e1000433

94. Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: Thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Archives of neurology*. 2004;61:1652-1661

95. Chung CM, Wang RY, Chen JW, Fann CS, Leu HB, Ho HY, Ting CT, Lin TH, Sheu SH, Tsai WC, Chen JH, Jong YS, Lin SJ, Chen YT, Pan WH. A genome-wide association study identifies new loci for ace activity: Potential implications for response to ace inhibitor. *Pharmacogenomics J*. 2010;10:537-544

96. Tulah AS, Parker SG, Moffatt MF, Wardlaw AJ, Connolly MJ, Sayers I. The role of alox5ap, lta4h and ltb4r polymorphisms in determining baseline lung function and copd susceptibility in uk smokers. *BMC medical genetics*. 2011;12:173

97. Chen J, Yang T, Yu H, Sun K, Shi Y, Song W, Bai Y, Wang X, Lou K, Song Y, Zhang Y, Hui R. A functional variant in the 3'-utr of angiopoietin-1 might reduce stroke risk by interfering with the binding efficiency of microrna 211. *Hum Mol Genet*. 2010;19:2524-2533

98. Chen J, Yu H, Song W, Sun K, Song Y, Lou K, Yang T, Zhang Y, Hui R. Angiopoietin-2 promoter haplotypes confer an increased risk of stroke in a chinese han population. *Clin Sci (Lond)*. 2009;117:387-395

99. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatou E, D'Alfonso S, Blackburn H, Martinelli Boneschi F, Liddle J, Harbo HF, Perez ML, Spurkland A, Waller MJ, Mycko MP, Ricketts M, Comabella M, Hammond N, Kockum I, McCann OT, Ban M, Whittaker P, Kemppinen A, Weston P, Hawkins C, Widaa S, Zajicek J, Dronov S, Robertson N, Bumpstead SJ, Barcellos LF, Ravindrarajah R, Abraham R, Alfredsson L, Ardlie K, Aubin C, Baker A, Baker K, Baranzini SE, Bergamaschi L, Bergamaschi R, Bernstein A, Berthele A, Boggild M, Bradfield

JP, Brassat D, Broadley SA, Buck D, Butzkueven H, Capra R, Carroll WM, Cavalla P, Celius EG, Cepok S, Chiavacci R, Clerget-Darpoux F, Clysters K, Comi G, Cossburn M, Cournu-Rebeix I, Cox MB, Cozen W, Cree BA, Cross AH, Cusi D, Daly MJ, Davis E, de Bakker PI, Debouverie M, D'Hooghe M B, Dixon K, Dobosi R, Dubois B, Ellinghaus D, Elovaara I, Esposito F, Fontenille C, Foote S, Franke A, Galimberti D, Ghezzi A, Glessner J, Gomez R, Gout O, Graham C, Grant SF, Guerini FR, Hakonarson H, Hall P, Hamsten A, Hartung HP, Heard RN, Heath S, Hobart J, Hoshi M, Infante-Duarte C, Ingram G, Ingram W, Islam T, Jagodic M, Kabesch M, Kermode AG, Kilpatrick TJ, Kim C, Klopp N, Koivisto K, Larsson M, Lathrop M, Lechner-Scott JS, Leone MA, Leppa V, Liljedahl U, Bomfim IL, Lincoln RR, Link J, Liu J, Lorentzen AR, Lupoli S, Macciardi F, Mack T, Marriott M, Martinelli V, Mason D, McCauley JL, Mentch F, Mero IL, Mihalova T, Montalban X, Mottershead J, Myhr KM, Naldi P, Ollier W, Page A, Palotie A, Pelletier J, Piccio L, Pickersgill T, Piehl F, Pobywajlo S, Quach HL, Ramsay PP, Reunanen M, Reynolds R, Rioux JD, Rodegher M, Roesner S, Rubio JP, Ruckert IM, Salvetti M, Salvi E, Santaniello A, Schaefer CA, Schreiber S, Schulze C, Scott RJ, Sellebjerg F, Selmaj KW, Sexton D, Shen L, Simms-Acuna B, Skidmore S, Sleiman PM, Smestad C, Sorensen PS, Sondergaard HB, Stankovich J, Strange RC, Sulonen AM, Sundqvist E, Syvanen AC, Taddeo F, Taylor B, Blackwell JM, Tienari P, Bramon E, Tourbah A, Brown MA, Tronczynska E, Casas JP, Tubridy N, Corvin A, Vickery J, Jankowski J, Villoslada P, Markus HS, Wang K, Mathew CG, Wason J, Palmer CN, Wichmann HE, Plomin R, Willoughby E, Rautanen A, Winkelmann J, Wittig M, Trembath RC, Yaouanq J, Viswanathan AC, Zhang H, Wood NW, Zuvich R, Deloukas P, Langford C, Duncanson A, Oksenberg JR, Pericak-Vance MA, Haines JL, Olsson T, Hillert J, Ivinson AJ, De Jager PL, Peltonen L, Stewart GJ, Hafler DA, Hauser SL, McVean G, Donnelly P, Compston A. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *NATURE*. 2011;476:214-219

100. Sudlow C, Martinez Gonzalez NA, Kim J, Clark C. Does apolipoprotein e genotype influence the risk of ischemic stroke, intracerebral hemorrhage, or subarachnoid hemorrhage? Systematic review and meta-analyses of 31 studies among 5961 cases and 17,965 controls. *Stroke*. 2006;37:364-370

101. Wang Q, Ding H, Tang JR, Zhang L, Xu YJ, Yan JT, Wang W, Hui RT, Wang CY, Wang DW. C-reactive protein polymorphisms and genetic susceptibility to ischemic stroke and hemorrhagic stroke in the chinese han population. *Acta pharmacologica Sinica*. 2009;30:291-298

102. Lange LA, Carlson CS, Hindorff LA, Lange EM, Walston J, Durda JP, Cushman M, Bis JC, Zeng D, Lin D, Kuller LH, Nickerson DA, Psaty BM, Tracy RP, Reiner AP. Association of polymorphisms in the crp gene with circulating c-reactive protein levels and cardiovascular events. *JAMA*. 2006;296:2703-2711

103. Coddington CC, Grow DR, Ahmed MS, Toner JP, Cook E, Diamond MP. Gonadotropin-releasing hormone agonist pretreatment did not decrease postoperative adhesion formation after abdominal myomectomy in a randomized control trial. *Fertility and sterility*. 2009;91:1909-1913

104. Ding H, Cui G, Zhang L, Xu Y, Bao X, Tu Y, Wu B, Wang Q, Hui R, Wang W, Dackor RT, Kissling GE, Zeldin DC, Wang DW. Association of common variants of cyp4a11 and cyp4f2 with stroke in the han chinese population. *Pharmacogenet Genomics*. 2010;20:187-194

105. Major JM, Yu K, Wheeler W, Zhang H, Cornelis MC, Wright ME, Yeager M, Snyder K, Weinstein SJ, Mondul A, Eliassen H, Purdue M, Hazra A, McCarty CA, Hendrickson S, Virtamo J, Hunter D, Chanock S, Kraft P, Albanes D. Genome-wide association study identifies common variants associated with circulating vitamin e levels. *Hum Mol Genet*. 2011;20:3876-3883

106. Tu Y, Cui G, Xu Y, Bao X, Wang X, Wang DW. Genetic polymorphism of cyp11b2 gene and stroke in the han chinese population and a meta-analysis. *Pharmacogenet Genomics*. 2011;21:115-120

107. Munshi A, Sharma V, Kaul S, Rajeshwar K, Babu MS, Shafi G, Anila AN, Balakrishna N, Alladi S, Jyothy A. Association of the -344c/t aldosterone synthase (cyp11b2) gene variant with hypertension and stroke. *J Neurol Sci*. 2010;296:34-38

108. Ding H, Wu B, Wang H, Lu Z, Yan J, Wang X, Shaffer JR, Hui R, Wang DW. A novel loss-of-function ddah1 promoter polymorphism is associated with increased susceptibility to thrombosis stroke and coronary heart disease. *Circ Res*. 2010;106:1145-1152

109. Tao HM, Chen GZ. Endothelial no synthase gene polymorphisms and risk of ischemic stroke: A meta-analysis. *Neurosci Res*. 2009;64:311-316

110. Chen H, Chu H, Shi Y, Bhuyan SS, Li J, Liu SR, Yang J. Association between endothelial nitric oxide synthase polymorphisms and atrial fibrillation: A meta-analysis. *Journal of cardiovascular translational research*. 2012;5:528-534

111. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, Cohen W, Oudot-Mellakh T, Antoni G, Alessi MC, Zelenika D, Cambien F, Tiret L, Bertrand M, Dupuy AM, Letenneur L, Lathrop M, Emmerich J, Amouyel P, Tregouet DA, Morange PE. Genetics of venous thrombosis: Insights from a new genome wide association study. *PLoS One*. 2011;6:e25581

112. Smith NL, Huffman JE, Strachan DP, Huang J, Dehghan A, Trompet S, Lopez LM, Shin SY, Baumert J, Vitart V, Bis JC, Wild SH, Rumley A, Yang Q, Uitterlinden AG, Stott DJ, Davies G, Carter AM, Thorand B, Polasek O, McKnight B, Campbell H, Rudnicka AR, Chen MH, Buckley BM, Harris SE, Peters A, Pulanic D, Lumley T, de Craen AJ, Liewald DC, Gieger C, Campbell S, Ford I, Gow AJ, Luciano M, Porteous DJ, Guo X, Sattar N, Tenesa A, Cushman M, Slagboom PE, Visscher PM, Spector TD, Illig T, Rudan I, Bovill EG, Wright AF, McArdle WL, Tofler G, Hofman A, Westendorp RG, Starr JM, Grant PJ, Karakas M, Hastie ND, Psaty BM, Wilson JF, Lowe GD, O'Donnell CJ, Witteman JC, Jukema JW, Deary IJ, Soranzo N, Koenig W, Hayward C. Genetic predictors of fibrin d-dimer levels in healthy adults. *Circulation*. 2011;123:1864-1872

113. Chen XC, Xu MT, Zhou W, Han CL, Chen WQ. A meta-analysis of beta-fibrinogen gene-455g/a polymorphism and plasma fibrinogen level in chinese cerebral infarction patients. *Biomedical and environmental sciences : BES*. 2007;20:366-372

114. Danik JS, Pare G, Chasman DI, Zee RY, Kwiatkowski DJ, Parker A, Miletich JP, Ridker PM. Novel loci, including those related to crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women: The women's genome health study. *Circ Cardiovasc Genet*. 2009;2:134-141

115. Lovely RS, Yang Q, Massaro JM, Wang J, D'Agostino RB, Sr., O'Donnell CJ, Shannon J, Farrell DH. Assessment of genetic determinants of the association of gamma' fibrinogen in relation to cardiovascular disease. *Arterioscler Thromb Vasc Biol*. 2011;31:2345-2352

116. Maguire JM, Thakkinstian A, Sturm J, Levi C, Lincz L, Parsons M, Whyte S, Attia J. Polymorphisms in platelet glycoprotein 1balpha and factor vii and risk of ischemic stroke: A meta-analysis. *Stroke*. 2008;39:1710-1716

117. Bentley P, Peck G, Smeeth L, Whittaker J, Sharma P. Causal relationship of susceptibility genes to ischemic stroke: Comparison to ischemic heart disease and biochemical determinants. *PLoS One*. 2010;5:e9136

118. Cornere B. Vibrio parahaemolyticus food poisoning. *N Z Med J*. 1978;87:63

119. Tong Y, Wang Z, Geng Y, Liu J, Zhang R, Lin Q, Li X, Huang D, Gao S, Hu D, Li Y, Cheng J, Lu Z. The association of functional polymorphisms of il-6 gene promoter with ischemic stroke: Analysis in two chinese populations. *Biochem Biophys Res Commun*. 2010;391:481-485

120. Fornage M, Chiang YA, O'Meara ES, Psaty BM, Reiner AP, Siscovick DS, Tracy RP, Longstreth WT, Jr. Biomarkers of inflammation and mri-defined small vessel disease of the brain: The cardiovascular health study. *Stroke*. 2008;39:1952-1959

121. Bevan S, Dichgans M, Wiechmann HE, Gschwendtner A, Meitinger T, Markus HS. Genetic variation in members of the leukotriene biosynthesis pathway confer an increased risk of ischemic stroke: A replication study in two independent populations. *Stroke*. 2008;39:1109-1114

122. Freiberg JJ, Dahl M, Tybjaerg-Hansen A, Grande P, Nordestgaard BG. Leukotriene c4 synthase and ischemic cardiovascular disease and obstructive pulmonary disease in 13,000 individuals. *Journal of molecular and cellular cardiology*. 2009;46:579-586

123. Pare G, Chasman DI, Parker AN, Zee RR, Malarstig A, Seedorf U, Collins R, Watkins H, Hamsten A, Miletich JP, Ridker PM. Novel associations of cps1, mut, nox4, and dpep1 with plasma homocysteine in a healthy population: A genome-wide evaluation of 13 974 participants in the women's genome health study. *Circ Cardiovasc Genet*. 2009;2:142-150

124. Kim NS, Ko MM, Cha MH, Oh SM, Bang OS. Age and sex dependent genetic effects of neuropeptide y promoter polymorphism on susceptibility to ischemic stroke in koreans. *Clin Chim Acta*. 2010;411:1243-1247

125. Attia J, Thakkinstian A, Wang Y, Lincz L, Parsons M, Sturm J, McGettigan P, Scott R, Meldrum C, Levi C. The pai-1 4g/5g gene polymorphism and ischemic stroke: An association study and meta-analysis. *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association*. 2007;16:173-179

126. Dahabreh IJ, Kitsios GD, Kent DM, Trikalinos TA. Paraoxonase 1 polymorphisms and ischemic stroke risk: A systematic review and meta-analysis. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2010;12:606-615

127. Anderson CD, Biffi A, Rost NS, Cortellini L, Furie KL, Rosand J. Chromosome 9p21 in ischemic stroke: Population structure and meta-analysis. *Stroke*. 2010;41:1123-1131

128. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR. The gene encoding phosphodiesterase 4d confers risk of ischemic stroke. *Nat Genet*. 2003;35:131-138

129. Sun Y, Huang Y, Chen X, Liu Y, Lu X, Shi Y, Tang W, Yang J, Chen W, Zhao X, Gao L, Li S, Feng G, He L. Association between the pde4d gene and ischaemic stroke in the chinese han population. *Clin Sci (Lond)*. 2009;117:265-272

130. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB, Willis-Owen SA, Klanderman B, Lasky-Su J, Lazarus R, Murphy AJ, Soto-Quiros ME, Avila L, Beaty T, Mathias RA, Ruczinski I, Barnes KC, Celedon JC, Cookson WO, Gauderman WJ, Gilliland FD, Hakonarson H, Lange C, Moffatt MF, O'Connor GT, Raby BA, Silverman EK, Weiss ST. Genome-wide association analysis identifies pde4d as an asthma-susceptibility gene. *Am J Hum Genet*. 2009;84:581-593

131. Gottlieb DJ, O'Connor GT, Wilk JB. Genome-wide association of sleep and circadian phenotypes. *BMC medical genetics*. 2007;8 Suppl 1:S9

132. Wu C, Hu Z, He Z, Jia W, Wang F, Zhou Y, Liu Z, Zhan Q, Liu Y, Yu D, Zhai K, Chang J, Qiao Y, Jin G, Shen Y, Guo C, Fu J, Miao X, Tan W, Shen H, Ke Y, Zeng Y, Wu T, Lin D. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in chinese populations. *Nat Genet*. 2011;43:679-684

133. Dahlberg J, Smith G, Norrving B, Nilsson P, Hedblad B, Engstrom G, Lovkvist H, Carlson J, Lindgren A, Melander O. Genetic variants in serum and glucocortocoid regulated kinase 1, a regulator of the epithelial sodium channel, are associated with ischaemic stroke. *J Hypertens*. 2011;29:884-889

134. Munshi A, Rajeshwar K, Kaul S, Al-Hazzani A, Alshatwi AA, Shafi G, Balakrishna N, Jyothy A. Association of tumor necrosis factor-alpha and matrix metalloproteinase-3 gene variants with stroke. *Eur J Neurol*. 2011;18:1053-1059

135. Tong Y, Geng Y, Xu J, Wang Z, Zhang Y, Lin L, Zhang R, Deng P, Li Y, Hou W, Chai Y, Mason KA, Lu Z, Cheng J. The role of functional polymorphisms of the tnf-alpha gene promoter in the risk of ischemic stroke in chinese han and uyghur populations: Two case-control studies. *Clin Chim Acta*. 2010;411:1291-1295

136. Boraska V, Rayner NW, Groves CJ, Frayling TM, Diakite M, Rockett KA, Kwiatkowski DP, Day-Williams AG, McCarthy MI, Zeggini E. Large-scale association analysis of tnf/lta gene region polymorphisms in type 2 diabetes. *BMC medical genetics*. 2010;11:69

137. Wang Y, Zhang W, Zhang Y, Yang Y, Sun L, Hu S, Chen J, Zhang C, Zheng Y, Zhen Y, Sun K, Fu C, Yang T, Wang J, Sun J, Wu H, Glasgow WC, Hui R. Vkorc1 haplotypes are associated

with arterial vascular diseases (stroke, coronary heart disease, and aortic dissection). *Circulation*. 2006;113:1615-1621

138. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, Debette S, Lumley T, Folsom AR, van den Herik EG, Bos MJ, Beiser A, Cushman M, Launer LJ, Shahar E, Struchalin M, Du Y, Glazer NL, Rosamond WD, Rivadeneira F, Kelly-Hayes M, Lopez OL, Coresh J, Hofman A, DeCarli C, Heckbert SR, Koudstaal PJ, Yang Q, Smith NL, Kase CS, Rice K, Haritunians T, Roks G, de Kort PL, Taylor KD, de Lau LM, Oostra BA, Uitterlinden AG, Rotter JI, Boerwinkle E, Psaty BM, Mosley TH, van Duijn CM, Breteler MM, Longstreth WT, Jr., Wolf PA. Genomewide association studies of stroke. *N Engl J Med*. 2009;360:1718-1728

139. Matsushita T, Umeno J, Hirakawa Y, Yonemoto K, Ashikawa K, Amitani H, Ninomiya T, Hata J, Doi Y, Kitazono T, Iida M, Nakamura Y, Kiyohara Y, Kubo M. Association study of the polymorphisms on chromosome 12p13 with atherothrombotic stroke in the japanese population. *Journal of human genetics*. 2010;55:473-476

140. Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, Matsushita T, Yamazaki K, Ohnishi Y, Saito S, Kitazono T, Ibayashi S, Sueishi K, Iida M, Nakamura Y, Kiyohara Y. A nonsynonymous snp in prkch (protein kinase c eta) increases the risk of cerebral infarction. *Nat Genet*. 2007;39:212-217

141. Wu L, Shen Y, Liu X, Ma X, Xi B, Mi J, Lindpaintner K, Tan X, Wang X. The 1425g/a snp in prkch is associated with ischemic stroke and cerebral hemorrhage in a chinese population. *Stroke*. 2009;40:2973-2976

142. Garcia-Berrocoso T, Fernandez-Cadenas I, Delgado P, Rosell A, Montaner J. Blood biomarkers in cardioembolic stroke. *Current cardiology reviews*. 2010;6:194-201

143. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, Fornage M, Ikram MA, Malik R, Bevan S, Thorsteinsdottir U, Nalls MA, Longstreth W, Wiggins KL, Yadav S, Parati EA, Destefano AL, Worrall BB, Kittner SJ, Khan MS, Reiner AP, Helgadottir A, Achterberg S, Fernandez-Cadenas I, Abboud S, Schmidt R, Walters M, Chen WM, Ringelstein EB, O'Donnell M, Ho WK, Pera J, Lemmens R, Norrving B, Higgins P, Benn M, Sale M, Kuhlenbaumer G, Doney AS, Vicente AM, Delavaran H, Algra A, Davies G, Oliveira SA, Palmer CN, Deary I, Schmidt H, Pandolfo M, Montaner J, Carty C, de Bakker PI, Kostulas K, Ferro JM, van Zuydam NR, Valdimarsson E, Nordestgaard BG, Lindgren A, Thijs V, Slowik A, Saleheen D, Pare G, Berger K, Thorleifsson G, Hofman A, Mosley TH, Mitchell BD, Furie K, Clarke R, Levi C, Seshadri S, Gschwendtner A, Boncoraglio GB, Sharma P, Bis JC, Gretarsdottir S, Psaty BM, Rothwell PM, Rosand J, Meschia JF, Stefansson K, Dichgans M, Markus HS. Genetic risk factors for ischaemic stroke and its subtypes (the metastroke collaboration): A meta-analysis of genome-wide association studies. *Lancet neurology*. 2012

144. Gudbjartsson DF, Holm H, Gretarsdottir S, Thorleifsson G, Walters GB, Thorgeirsson G, Gulcher J, Mathiesen EB, Njolstad I, Nyrnes A, Wilsgaard T, Hald EM, Hveem K, Stoltenberg C, Kucera G, Stubblefield T, Carter S, Roden D, Ng MC, Baum L, So WY, Wong KS, Chan JC, Gieger C, Wichmann HE, Gschwendtner A, Dichgans M, Kuhlenbaumer G, Berger K, Ringelstein EB, Bevan S, Markus HS, Kostulas K, Hillert J, Sveinbjornsdottir S, Valdimarsson EM, Lochen ML, Ma RC, Darbar D, Kong A, Arnar DO, Thorsteinsdottir U, Stefansson K. A

sequence variant in zfhx3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet*. 2009;41:876-878

145. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, Jackson CA, Traylor M, Strange A, Su Z, Band G, Syme PD, Malik R, Pera J, Norrving B, Lemmens R, Freeman C, Schanz R, James T, Poole D, Murphy L, Segal H, Cortellini L, Cheng YC, Woo D, Nalls MA, Muller-Myhsok B, Meisinger C, Seedorf U, Ross-Adams H, Boonen S, Wloch-Kopec D, Valant V, Slark J, Furie K, Delavaran H, Langford C, Deloukas P, Edkins S, Hunt S, Gray E, Dronov S, Peltonen L, Gretarsdottir S, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Boncoraglio GB, Parati EA, Attia J, Holliday E, Levi C, Franzosi MG, Goel A, Helgadottir A, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin A, Duncanson A, Jankowski J, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Trembath RC, Viswanathan AC, Wood NW, Worrall BB, Kittner SJ, Mitchell BD, Kissela B, Meschia JF, Thijs V, Lindgren A, Macleod MJ, Slowik A, Walters M, Rosand J, Sharma P, Farrall M, Sudlow CL, Rothwell PM, Dichgans M, Donnelly P, Markus HS. Genome-wide association study identifies a variant in hdac9 associated with large vessel ischemic stroke. *Nat Genet*. 2012;44:328-333

146. Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD. Genome-wide association to body mass index and waist circumference: The framingham heart study 100k project. *BMC medical genetics*. 2007;8 Suppl 1:S18

147. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang SJ, Vasan RS, Mitchell GF. Framingham heart study 100k project: Genome-wide associations for blood pressure and arterial stiffness. *BMC medical genetics*. 2007;8 Suppl 1:S3

148. Criqui MH. Peripheral arterial disease--epidemiological aspects. *Vasc Med*. 2001;6:3-7

149. Hiatt WR, Hoag S, Hamman RF. Effect of diagnostic criteria on the prevalence of peripheral arterial disease. The san luis valley diabetes study. *Circulation*. 1995;91:1472-1479

150. Gregg EW, Sorlie P, Paulose-Ram R, Gu Q, Eberhardt MS, Wolz M, Burt V, Curtin L, Engelgau M, Geiss L. Prevalence of lower-extremity disease in the us adult population >=40 years of age with and without diabetes: 1999-2000 national health and nutrition examination survey. *Diabetes Care*. 2004;27:1591-1597

151. Fowkes FGR. *Epidemiology of peripheral vascular disease*. London ; New York: Springer-Verlag; 1991.

152. Criqui MH, Langer RD, Fronek A, Feigelson HS, Klauber MR, McCann TJ, Browner D. Mortality over a period of 10 years in patients with peripheral arterial disease. *New England Journal of Medicine*. 1992;326:381-386

153. Hertzer NR, Beven EG, Young JR, O'Hara PJ, Ruschhaupt WF, 3rd, Graor RA, Dewolfe VG, Maljovec LC. Coronary artery disease in peripheral vascular patients. A classification of 1000 coronary angiograms and results of surgical management. *Annals of surgery*. 1984;199:223-233

154. Dormandy J, Mahir M, Ascady G, Balsano F, De Leeuw P, Blombery P, Bousser MG, Clement D, Coffman J, Deutshinoff A, et al. Fate of the patient with chronic leg ischaemia. A review article. *The Journal of cardiovascular surgery*. 1989;30:50-57

155. Aronow WS, Ahn C. Prevalence of coexistence of coronary artery disease, peripheral arterial disease, and atherothrombotic brain infarction in men and women > or = 62 years of age. *Am J Cardiol*. 1994;74:64-65

156. Association AD. Peripheral arterial disease in people with diabetes. *Diabetes Care*. 2003;26:3333-3341

157. White CJ, Gray WA. Endovascular therapies for peripheral arterial disease: An evidence-based review. *Circulation*. 2007;116:2203-2215

158. Murabito JM, Guo CY, Fox CS, D'Agostino RB. Heritability of the ankle-brachial index: The framingham offspring study. *Am J Epidemiol*. 2006;164:963-968

159. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsater A, Flex A, Aben KK, de Vegt F, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *NATURE*. 2008;452:638-642

160. Murabito JM, White CC, Kavousi M, Sun YV, Feitosa MF, Nambi V, Lamina C, Schillert A, Coassin S, Bis JC, Broer L, Crawford DC, Franceschini N, Frikke-Schmidt R, Haun M, Holewijn S, Huffman JE, Hwang SJ, Kiechl S, Kollerits B, Montasser ME, Nolte IM, Rudock ME, Senft A, Teumer A, van der Harst P, Vitart V, Waite LL, Wood AR, Wassel CL, Absher DM, Allison MA, Amin N, Arnold A, Asselbergs FW, Aulchenko Y, Bandinelli S, Barbalic M, Boban M, Brown-Gentry K, Couper DJ, Criqui MH, Dehghan A, den Heijer M, Dieplinger B, Ding J, Dorr M, Espinola-Klein C, Felix SB, Ferrucci L, Folsom AR, Fraedrich G, Gibson Q, Goodloe R, Gunjaca G, Haltmayer M, Heiss G, Hofman A, Kieback A, Kiemeney LA, Kolcic I, Kullo IJ, Kritchevsky SB, Lackner KJ, Li X, Lieb W, Lohman K, Meisinger C, Melzer D, Mohler ER, 3rd, Mudnic I, Mueller T, Navis G, Oberhollenzer F, Olin JW, O'Connell J, O'Donnell CJ, Palmas W, Penninx BW, Petersmann A, Polasek O, Psaty BM, Rantner B, Rice K, Rivadeneira F, Rotter JI, Seldenrijk A, Stadler M, Summerer M, Tanaka T, Tybjaerg-Hansen A, Uitterlinden AG, van Gilst WH, Vermeulen SH, Wild SH, Wild PS, Willeit J, Zeller T, Zemunik T, Zgaga L, Assimes TL, Blankenberg S, Boerwinkle E, Campbell H, Cooke JP, de Graaf J, Herrington D, Kardia SL, Mitchell BD, Murray A, Munzel T, Newman AB, Oostra BA, Rudan I, Shuldiner AR, Snieder H, van Duijn CM, Volker U, Wright AF, Wichmann HE, Wilson JF, Witteman JC, Liu Y, Hayward C, Borecki IB, Ziegler A, North KE, Cupples LA, Kronenberg F. Association between chromosome 9p21 variants and the ankle-brachial index identified by a meta-analysis of 21 genome-wide association studies. *Circ Cardiovasc Genet*. 2012;5:100-112

161. Smith JG, Melander O, Lovkvist H, Hedblad B, Engstrom G, Nilsson P, Carlson J, Berglund G, Norrving B, Lindgren A. Common genetic variants on chromosome 9p21 confers risk of ischemic stroke: A large-scale genetic association study. *Circ Cardiovasc Genet*. 2009;2:159-164

162. Hopewell JC, Clarke R, Parish S, Armitage J, Lathrop M, Hager J, Collins R. Lipoprotein(a) genetic variants associated with coronary and peripheral vascular disease but not with stroke risk in the heart protection study. *Circ Cardiovasc Genet*. 2011;4:68-73

163. Xi B, Shen Y, Yan Y, Mi J. Association of polymorphisms in the agt gene with essential hypertension in the chinese population. *Journal of the renin-angiotensin-aldosterone system : JRAAS*. 2012;13:282-288

164. Murabito JM, Evans JC, Nieto K, Larson MG, Levy D, Wilson PW. Prevalence and clinical correlates of peripheral arterial disease in the framingham offspring study. *Am Heart J*. 2002;143:961-965

165. Nitschke Y, Hartmann S, Torsello G, Horstmann R, Seifarth H, Weissen-Plenz G, Rutsch F. Expression of npp1 is regulated during atheromatous plaque calcification. *Journal of cellular and molecular medicine*. 2011;15:220-231

166. Cote N, El Husseini D, Pepin A, Guauque-Olarte S, Ducharme V, Bouchard-Cannon P, Audet A, Fournier D, Gaudreault N, Derbali H, McKee MD, Simard C, Despres JP, Pibarot P, Bosse Y, Mathieu P. Atp acts as a survival signal and prevents the mineralization of aortic valve. *Journal of molecular and cellular cardiology*. 2012;52:1191-1202

167. Basar Y, Salmayenli N, Aksoy M, Seckin S, Aydin M, Ozkok E. Ace gene polymorphism in peripheral vascular disease. *Horm Metab Res*. 2007;39:534-537

168. Consortium. TI-RMRAIRM. The interleukin-6 receptor as a target for prevention of coronary heart disease: A mendelian randomisation analysis. *Lancet*. 2012

169. Zintzaras E, Zdoukopoulos N. A field synopsis and meta-analysis of genetic association studies in peripheral arterial disease: The cumagas-pad database. *Am J Epidemiol*. 2009;170:1-11

170. Koriyama H, Nakagami H, Katsuya T, Sugimoto K, Yamashita H, Takami Y, Maeda S, Kubo M, Takahashi A, Nakamura Y, Ogihara T, Rakugi H, Kaneda Y, Morishita R. Identification of evidence suggestive of an association with peripheral arterial disease at the osbpl10 locus by genome-wide investigation in the japanese population. *J Atheroscler Thromb*. 2010;17:1054-1062

171. Kamal M, Shaaban AM, Zhang L, Walker C, Gray S, Thakker N, Toomes C, Speirs V, Bell SM. Loss of csmd1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma. *Breast cancer research and treatment*. 2010;121:555-563

172. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, Bennett F, Bowden DW, Chakravarti A, Dreisbach A, Farlow DN, Folsom AR, Fornage M, Forrester T, Fox E, Haiman CA, Hartiala J, Harris TB, Hazen SL, Heckbert SR, Henderson BE, Hirschhorn JN, Keating BJ, Kritchevsky SB, Larkin E, Li M, Rudock ME, McKenzie CA, Meigs JB, Meng YA, Mosley TH, Newman AB, Newton-Cheh CH, Paltoo DN, Papanicolaou GJ, Patterson N, Post

WS, Psaty BM, Qasim AN, Qu L, Rader DJ, Redline S, Reilly MP, Reiner AP, Rich SS, Rotter JI, Liu Y, Shrader P, Siscovick DS, Tang WH, Taylor HA, Tracy RP, Vasan RS, Waters KM, Wilks R, Wilson JG, Fabsitz RR, Gabriel SB, Kathiresan S, Boerwinkle E. Genome-wide association study of coronary heart disease and its risk factors in 8,090 african americans: The nhlbi care project. *PLoS Genet*. 2011;7:e1001300

173. Eller P, Schgoer W, Mueller T, Tancevski I, Demetz E, Duwensee K, Ritsch A, Haltmayer M, Patsch JR. The k121q polymorphism of enpp1 and peripheral arterial disease. *Heart Vessels*. 2008;23:104-107

174. Yazdani-Biuki B, Krippl P, Brickmann K, Fuerst F, Langsenlehner U, Paulweber B, Pilger E, Wascher TC, Brezinschek HP, Renner W. The functional promoter polymorphism of the coagulation factor xii gene is not associated with peripheral arterial disease. *Angiology*. 2010;61:211-215

175. Klein RL, Hunter SJ, Jenkins AJ, Zheng D, Semler AJ, Clore J, Garvey WT. Fibrinogen is a marker for nephropathy and peripheral vascular disease in type 1 diabetes: Studies of plasma fibrinogen and fibrinogen gene polymorphism in the dcct/edic cohort. *Diabetes Care*. 2003;26:1439-1448

176. Khandanpour N, Willis G, Meyer FJ, Armon MP, Loke YK, Wright AJ, Finglas PM, Jennings BA. Peripheral arterial disease and methylenetetrahydrofolate reductase (mthfr) c677t mutations: A case-control study and meta-analysis. *J Vasc Surg*. 2009;49:711-718

177. Schgoer W, Eller P, Mueller T, Tancevski I, Wehinger A, Ulmer H, Sandhofer A, Ritsch A, Haltmayer M, Patsch JR. The mtp -493tt genotype is associated with peripheral arterial disease: Results from the linz peripheral arterial disease (lipad) study. *Clinical biochemistry*. 2008;41:712-716

178. El-Koofy NM, El-Karaksy HM, Mandour IM, Anwar GM, El-Raziky MS, El-Hennawy AM. Genetic polymorphisms in non-alcoholic fatty liver disease in obese egyptian children. *Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association*. 2011;17:265-270

179. Sticchi E, Sofi F, Romagnuolo I, Pratesi G, Pulli R, Pratesi C, Abbate R, Fatini C. Enos and ace genes influence peripheral arterial disease predisposition in smokers. *J Vasc Surg*. 2010;52:97-102 e101

180. Kullo IJ, Greene MT, Boerwinkle E, Chu J, Turner ST, Kardia SL. Association of polymorphisms in nos3 with the ankle-brachial index in hypertensive adults. *Atherosclerosis*. 2008;196:905-912

181. Folsom AR, Peacock JM, Boerwinkle E. Variation in pcsk9, low ldl cholesterol, and risk of peripheral arterial disease. *Atherosclerosis*. 2009;202:211-215

182. Catalano M, Cortelazzo A, Santi R, Contino L, Demicheli M, Yilmaz Y, Zorzetto M, Campo I, Lanati N, Emanuele E. The pro12ala polymorphism of peroxisome proliferator-activated receptor-gamma2 gene is associated with plasma levels of soluble rage (receptor for advanced glycation endproducts) and the presence of peripheral arterial disease. *Clinical biochemistry*. 2008;41:981-985

183. Matsuo T, Nakata Y, Katayama Y, Iemitsu M, Maeda S, Okura T, Kim MK, Ohkubo H, Hotta K, Tanaka K. Pparg genotype accounts for part of individual variation in body weight reduction in response to calorie restriction. *Obesity (Silver Spring)*. 2009;17:1924-1931

184. Florez JC, Jablonski KA, Sun MW, Bayley N, Kahn SE, Shamoon H, Hamman RF, Knowler WC, Nathan DM, Altshuler D. Effects of the type 2 diabetes-associated pparg p12a polymorphism on progression to diabetes and response to troglitazone. *J Clin Endocrinol Metab*. 2007;92:1502-1509

185. Qu A, Shah YM, Manna SK, Gonzalez FJ. Disruption of endothelial peroxisome proliferator-activated receptor gamma accelerates diet-induced atherogenesis in ldl receptor-null mice. *Arterioscler Thromb Vasc Biol*. 2012;32:65-73

186. Kim RJ, Becker RC. Association between factor v leiden, prothrombin g20210a, and methylenetetrahydrofolate reductase c677t mutations and events of the arterial circulatory system: A meta-analysis of published studies. *Am Heart J*. 2003;146:948-957

187. Gordon T, Kannel WB. Multiple risk functions for predicting coronary heart disease: The concept, accuracy, and application. *Am Heart J*. 1982;103:1031-1039

188. Kannel WB, McGee DL. Diabetes and glucose tolerance as risk factors for cardiovascular disease: The framingham study. *Diabetes Care*. 1979;2:120-126

189. Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, Jacobs DR, Jr., Bangdiwala S, Tyroler HA. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective american studies. *Circulation*. 1989;79:8-15

190. Levitzky YS, Pencina MJ, D'Agostino RB, Meigs JB, Murabito JM, Vasan RS, Fox CS. Impact of impaired fasting glucose on cardiovascular disease: The framingham heart study. *J Am Coll Cardiol*. 2008;51:264-270

191. Barrett J, Clayton D, Concannon P, Akolkar B, Cooper J, Erlich H, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth D, Stevens H, Todd J, Walker N, Rich S, Type 1 Diabetes Genetics C. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature genetics*. 2009;41:703-710

192. Shu XO, Long J, Cai Q, Qi L, Xiang YB, Cho YS, Tai ES, Li X, Lin X, Chow WH, Go MJ, Seielstad M, Bao W, Li H, Cornelis MC, Yu K, Wen W, Shi J, Han BG, Sim XL, Liu L, Qi Q, Kim HL, Ng DP, Lee JY, Kim YJ, Li C, Gao YT, Zheng W, Hu FB. Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet*. 2010;6

193. Yamauchi T, Hara K, Maeda S, Yasuda K, Takahashi A, Horikoshi M, Nakamura M, Fujita H, Grarup N, Cauchi S, Ng DP, Ma RC, Tsunoda T, Kubo M, Watada H, Maegawa H, Okada-Iwabu M, Iwabu M, Shojima N, Shin HD, Andersen G, Witte DR, Jorgensen T, Lauritzen T, Sandbaek A, Hansen T, Ohshige T, Omori S, Saito I, Kaku K, Hirose H, So WY, Beury D, Chan JC, Park KS, Tai ES, Ito C, Tanaka Y, Kashiwagi A, Kawamori R, Kasuga M, Froguel P, Pedersen O, Kamatani N, Nakamura Y, Kadowaki T. A genome-wide association study in the japanese population identifies susceptibility loci for type 2 diabetes at ube2e2 and c2cd4a-c2cd4b. *Nat Genet*. 2010;42:864-868

194. Consortium TaG. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42:441-447

195. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Magi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tonjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Doring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Jarvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, McCarthy MI, Peltonen L, Thorsteinsdottir U, Stefansson K. Sequence variants at chrnb3-chrna6 and cyp2a6 affect smoking behavior. *Nat Genet*. 2010;42:448-453

196. Zittermann A, Schleithoff SS, Koerfer R. Putting cardiovascular disease and vitamin d insufficiency into perspective. *Br J Nutr*. 2005;94:483-492

197. Autier P, Gandini S. Vitamin d supplementation and total mortality: A meta-analysis of randomized controlled trials. *Arch Intern Med*. 2007;167:1730-1737

198. Grimes DS, Hindle E, Dyer T. Respiratory infection and coronary heart disease: Progression of a paradigm. *QJM : monthly journal of the Association of Physicians*. 2000;93:375-383

199. Vieth R, Ladak Y, Walfish PG. Age-related changes in the 25-hydroxyvitamin d versus parathyroid hormone relationship suggest a different reason why older adults require more vitamin d. *J Clin Endocrinol Metab*. 2003;88:185-191

200. Zittermann A. Vitamin d in preventive medicine: Are we ignoring the evidence? *Br J Nutr*. 2003;89:552-572

201. Feldman D, Glorieux FH, Pike JW. *Vitamin d*. San Diego: Academic Press; 1997.

202. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The ncbi dbgap database of genotypes and phenotypes. *Nat Genet*. 2007;39:1181-1186

203. Holick MF, Schnoes HK, DeLuca HF, Suda T, Cousins RJ. Isolation and identification of 1,25-dihydroxycholecalciferol. A metabolite of vitamin d active in intestine. *Biochemistry*. 1971;10:2799-2804

204. Yang ES, Burnstein KL. Vitamin d inhibits g1 to s progression in lncap prostate cancer cells through p27kip1 stabilization and cdk2 mislocalization to the cytoplasm. *J Biol Chem*. 2003;278:46862-46868

205. Zehnder D, Quinkler M, Eardley KS, Bland R, Lepenies J, Hughes SV, Raymond NT, Howie AJ, Cockwell P, Stewart PM, Hewison M. Reduction of the vitamin d hormonal system in kidney disease is associated with increased renal inflammation. *Kidney Int*. 2008;74:1343-1353

206. Pfeifer M, Begerow B, Minne HW. Vitamin d and muscle function. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2002;13:187-194

207. Haussler M, Whitfield G, Haussler C, Hsieh J, Thompson P, Selznick S, Dominguez C, Jurutka P. The nuclear vitamin d receptor: Biological and molecular regulatory properties revealed. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 1998;13:325-349

208. Davies MR, Hruska KA. Pathophysiological mechanisms of vascular calcification in end-stage renal disease. *Kidney Int*. 2001;60:472-479

209. Fraser JD, Otawara Y, Price PA. 1,25-dihydroxyvitamin d3 stimulates the synthesis of matrix gamma-carboxyglutamic acid protein by osteosarcoma cells. Mutually exclusive expression of vitamin k-dependent bone proteins by clonal osteoblastic cell lines. *J Biol Chem*. 1988;263:911-916

210. Luo G, Ducy P, McKee MD, Pinero GJ, Loyer E, Behringer RR, Karsenty G. Spontaneous calcification of arteries and cartilage in mice lacking matrix gla protein. *NATURE*. 1997;386:78-81

211. Shanahan CM, Weissberg PL. Smooth muscle cell heterogeneity: Patterns of gene expression in vascular smooth muscle cells in vitro and in vivo. *Arterioscler Thromb Vasc Biol*. 1998;18:333-338

212. Schurgers LJ, Dissel PE, Spronk HM, Soute BA, Dhore CR, Cleutjens JP, Vermeer C. Role of vitamin k and vitamin k-dependent proteins in vascular calcification. *Zeitschrift fur Kardiologie*. 2001;90 Suppl 3:57-63

213. Muller K, Haahr PM, Diamant M, Rieneck K, Kharazmi A, Bendtzen K. 1,25-dihydroxyvitamin d3 inhibits cytokine production by human blood monocytes at the post-transcriptional level. *Cytokine*. 1992;4:506-512

214. Canning MO, Grotenhuis K, de Wit H, Ruwhof C, Drexhage HA. 1-alpha,25-dihydroxyvitamin d3 (1,25(oh)(2)d(3)) hampers the maturation of fully active immature dendritic cells from monocytes. *Eur J Endocrinol*. 2001;145:351-357

215. Mallat Z, Besnard S, Duriez M, Deleuze V, Emmanuel F, Bureau MF, Soubrier F, Esposito B, Duez H, Fievet C, Staels B, Duverger N, Scherman D, Tedgui A. Protective role of interleukin-10 in atherosclerosis. *Circ Res*. 1999;85:e17-24

216. Rostand SG, Drueke TB. Parathyroid hormone, vitamin d, and cardiovascular disease in chronic renal failure. *Kidney Int*. 1999;56:383-392

217. Li YC, Kong J, Wei M, Chen ZF, Liu SQ, Cao LP. 1,25-dihydroxyvitamin d(3) is a negative endocrine regulator of the renin-angiotensin system. *J Clin Invest*. 2002;110:229-238

218. Tiosano D, Schwartz Y, Braver Y, Hadash A, Gepstein V, Weisman Y, Lorber A. The renin-angiotensin system, blood pressure, and heart structure in patients with hereditary vitamin d-resistance rickets (hvdrr). *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2011;26:2252-2260

219. Kimura Y, Kawamura M, Owada M, Oshima T, Murooka M, Fujiwara T, Hiramori K. Effectiveness of 1,25-dihydroxyvitamin d supplementation on blood pressure reduction in a pseudohypoparathyroidism patient with high renin activity. *Intern Med*. 1999;38:31-35

220. Pfeifer M, Begerow B, Minne HW, Nachtigall D, Hansen C. Effects of a short-term vitamin d(3) and calcium supplementation on blood pressure and parathyroid hormone levels in elderly women. *J Clin Endocrinol Metab*. 2001;86:1633-1637

221. Rostand SG. Ultraviolet light may contribute to geographic and racial blood pressure differences. *Hypertension*. 1997;30:150-156

222. Krause R, Buhring M, Hopfenmuller W, Holick MF, Sharma AM. Ultraviolet b and blood pressure. *Lancet*. 1998;352:709-710

223. Sheehan NA, Meng S, Didelez V. Mendelian randomisation: A tool for assessing causality in observational epidemiology. *Methods Mol Biol*. 2011;713:153-166

224. Arruda V, von Zuben P, Chiaparini L, Annichino-Bizzacchi J, Costa F. The mutation ala677->val in the methylene tetrahydrofolate reductase gene: A risk factor for arterial disease and venous thrombosis. *Thromb Haemost*. 1997;77:818 - 821

225. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, MacDonald TM. The diabetes audit and research in tayside scotland (darts) study: Electronic record linkage to create a diabetes register. Darts/memo collaboration. *BMJ*. 1997;315:524-528

226. Doney AS, Dannfald J, Kimber CH, Donnelly LA, Pearson E, Morris AD, Palmer CN. The fto gene is associated with an atherogenic lipid profile and myocardial infarction in patients with type 2 diabetes: A genetics of diabetes audit and research study in tayside scotland (go-darts) study. *Circ Cardiovasc Genet*. 2009;2:255-259

227. Team RDC. R: A language and environment for statistical computing. 2011

228. Slinker BK, Glantz SA. Multiple linear regression: Accounting for multiple simultaneous determinants of a continuous dependent variable. *Circulation*. 2008;117:1732-1737

229. Cox DR, Oakes D. *Analysis of survival data*. London ; New York: Chapman and Hall; 1984.

230. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, Oppermann U, Dilthey A, Pirinen M, Stone MA, Appleton L, Moutsianas L, Leslie S, Wordsworth T, Kenna TJ, Karaderi T, Thomas GP, Ward MM, Weisman MH, Farrar C, Bradbury LA, Danoy P, Inman RD,

Maksymowych W, Gladman D, Rahman P, Morgan A, Marzo-Ortega H, Bowness P, Gaffney K, Gaston JS, Smith M, Bruges-Armas J, Couto AR, Sorrentino R, Paladini F, Ferreira MA, Xu H, Liu Y, Jiang L, Lopez-Larrea C, Diaz-Pena R, Lopez-Vazquez A, Zayats T, Band G, Bellenguez C, Blackburn H, Blackwell JM, Bramon E, Bumpstead SJ, Casas JP, Corvin A, Craddock N, Deloukas P, Dronov S, Duncanson A, Edkins S, Freeman C, Gillman M, Gray E, Gwilliam R, Hammond N, Hunt SE, Jankowski J, Jayakumar A, Langford C, Liddle J, Markus HS, Mathew CG, McCann OT, McCarthy MI, Palmer CN, Peltonen L, Plomin R, Potter SC, Rautanen A, Ravindrarajah R, Ricketts M, Samani N, Sawcer SJ, Strange A, Trembath RC, Viswanathan AC, Waller M, Weston P, Whittaker P, Widaa S, Wood NW, McVean G, Reveille JD, Wordsworth BP, Brown MA, Donnelly P. Interaction between erap1 and hla-b27 in ankylosing spondylitis implicates peptide handling in the mechanism for hla-b27 in disease susceptibility. *Nat Genet*. 2011;43:761-767

231. Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, Barton A, Band G, Bellenguez C, Bergboer JG, Blackwell JM, Bramon E, Bumpstead SJ, Casas JP, Cork MJ, Corvin A, Deloukas P, Dilthey A, Duncanson A, Edkins S, Estivill X, Fitzgerald O, Freeman C, Giardina E, Gray E, Hofer A, Huffmeier U, Hunt SE, Irvine AD, Jankowski J, Kirby B, Langford C, Lascorz J, Leman J, Leslie S, Mallbris L, Markus HS, Mathew CG, McLean WH, McManus R, Mossner R, Moutsianas L, Naluai AT, Nestle FO, Novelli G, Onoufriadis A, Palmer CN, Perricone C, Pirinen M, Plomin R, Potter SC, Pujol RM, Rautanen A, Riveira-Munoz E, Ryan AW, Salmhofer W, Samuelsson L, Sawcer SJ, Schalkwijk J, Smith CH, Stahle M, Su Z, Tazi-Ahnini R, Traupe H, Viswanathan AC, Warren RB, Weger W, Wolk K, Wood N, Worthington J, Young HS, Zeeuwen PL, Hayday A, Burden AD, Griffiths CE, Kere J, Reis A, McVean G, Evans DM, Brown MA, Barker JN, Peltonen L, Donnelly P, Trembath RC. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between hla-c and erap1. *Nat Genet*. 2010;42:985-990

232. Zhou K, Bellenguez C, Spencer CC, Bennett AJ, Coleman RL, Tavendale R, Hawley SA, Donnelly LA, Schofield C, Groves CJ, Burch L, Carr F, Strange A, Freeman C, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin A, Craddock N, Deloukas P, Dronov S, Duncanson A, Edkins S, Gray E, Hunt S, Jankowski J, Langford C, Markus HS, Mathew CG, Plomin R, Rautanen A, Sawcer SJ, Samani NJ, Trembath R, Viswanathan AC, Wood NW, Harries LW, Hattersley AT, Doney AS, Colhoun H, Morris AD, Sutherland C, Hardie DG, Peltonen L, McCarthy MI, Holman RR, Palmer CN, Donnelly P, Pearson ER. Common variants near atm are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet*. 2011;43:117-120

233. Voorham J, Haaijer-Ruskamp F, Stolk R, Wolffenbuttel B, Denig P, Group GItATDT. The influence of elevated cardiometabolic risk factor levels on treatment changes in type 2 diabetes. *Diabetes Care*. 2008;31:501 - 503

234. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, Burtt NP, Fuchsberger C, Li Y, Erdmann J, Frayling TM, Heid IM, Jackson AU, Johnson T, Kilpelainen TO, Lindgren CM, Morris AP, Prokopenko I, Randall JC, Saxena R, Soranzo N, Speliotes EK, Teslovich TM, Wheeler E, Maguire J, Parkin M, Potter S, Rayner NW, Robertson N, Stirrups K, Winckler W, Sanna S, Mulas A, Nagaraja R, Cucca F, Barroso I, Deloukas P, Loos RJ, Kathiresan S, Munroe PB, Newton-Cheh C, Pfeufer A, Samani NJ, Schunkert H, Hirschhorn JN, Altshuler D, McCarthy MI, Abecasis GR, Boehnke M. The metabochip, a custom genotyping array for

genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet*. 2012;8:e1002793

235. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575

236. Browning SR, Browning BL. Haplotype phasing: Existing methods and new developments. *Nat Rev Genet*. 2011;12:703-714

237. Delaneau O, Coulonges C, Zagury JF. Shape-it: New rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics*. 2008;9:540

238. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529

239. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nature methods*. 2011

240. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;68:978-989

241. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JRB, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney ASF, Consortium TWTCC, McCarthy MI, Hattersley AT. Replication of genome-wide association signals in uk samples reveals risk loci for type 2 diabetes. *Science*. 2007;316:1336-1341

242. Colin Cameron A, Windmeijer FAG. An r-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*. 1997;77:329-342

243. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*. 1982;143:29-36

244. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: From area under the roc curve to reclassification and beyond. *Stat Med*. 2008;27:157-172; discussion 207-112

245. de Bakker P, Ferreira M, Jia X, Neale B, Raychaudhuri S, Voight B. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Human molecular genetics*. 2008;17:8

246. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88:586-598

247. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, Speliotes EK, Thorleifsson G, Willer CJ, Herrera BM, Jackson AU, Lim N, Scheet P, Soranzo N, Amin N, Aulchenko YS, Chambers JC, Drong A, Luan J, Lyon HN, Rivadeneira F, Sanna S, Timpson NJ, Zillikens MC, Zhao JH, Almgren P, Bandinelli S, Bennett AJ, Bergman RN, Bonnycastle LL,

Bumpstead SJ, Chanock SJ, Cherkas L, Chines P, Coin L, Cooper C, Crawford G, Doering A, Dominiczak A, Doney AS, Ebrahim S, Elliott P, Erdos MR, Estrada K, Ferrucci L, Fischer G, Forouhi NG, Gieger C, Grallert H, Groves CJ, Grundy S, Guiducci C, Hadley D, Hamsten A, Havulinna AS, Hofman A, Holle R, Holloway JW, Illig T, Isomaa B, Jacobs LC, Jameson K, Jousilahti P, Karpe F, Kuusisto J, Laitinen J, Lathrop GM, Lawlor DA, Mangino M, McArdle WL, Meitinger T, Morken MA, Morris AP, Munroe P, Narisu N, Nordstrom A, Nordstrom P, Oostra BA, Palmer CN, Payne F, Peden JF, Prokopenko I, Renstrom F, Ruokonen A, Salomaa V, Sandhu MS, Scott LJ, Scuteri A, Silander K, Song K, Yuan X, Stringham HM, Swift AJ, Tuomi T, Uda M, Vollenweider P, Waeber G, Wallace C, Walters GB, Weedon MN, Witteman JC, Zhang C, Zhang W, Caulfield MJ, Collins FS, Davey Smith G, Day IN, Franks PW, Hattersley AT, Hu FB, Jarvelin MR, Kong A, Kooner JS, Laakso M, Lakatta E, Mooser V, Morris AD, Peltonen L, Samani NJ, Spector TD, Strachan DP, Tanaka T, Tuomilehto J, Uitterlinden AG, van Duijn CM, Wareham NJ, Hugh W, Waterworth DM, Boehnke M, Deloukas P, Groop L, Hunter DJ, Thorsteinsdottir U, Schlessinger D, Wichmann HE, Frayling TM, Abecasis GR, Hirschhorn JN, Loos RJ, Stefansson K, Mohlke KL, Barroso I, McCarthy MI. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet*. 2009;5:e1000508

248. Cooper HM, Hedges LV, Valentine JC. *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation; 2009.

249. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560

250. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or i2 index? *Psychological methods*. 2006;11:193-206

251. Altman DG, Bland JM. Interaction revisited: The difference between two estimates. *BMJ*. 2003;326:219

252. Magi R, Morris AP. Gwama: Software for genome-wide association meta-analysis. *BMC Bioinformatics*. 2010;11:288

253. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, Li R, Masys DR, Ritchie MD, Roden DM, Struewing JP, Wolf WA. The emerge network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC medical genomics*. 2011;4:13

254. Metspalu A. Estonian genome project--before the take-off and take-off. *Bioinformatics*. 2002;18 Suppl 2:S152

255. Zhou K, Donnelly L, Burch L, Tavendale R, Doney AS, Leese G, Hattersley AT, McCarthy MI, Morris AD, Lang CC, Palmer CN, Pearson ER. Loss-of-function cyp2c9 variants improve therapeutic response to sulfonylureas in type 2 diabetes: A go-darts study. *Clinical pharmacology and therapeutics*. 2010;87:52-56

256. Donnelly LA, Palmer CN, Whitley AL, Lang CC, Doney AS, Morris AD, Donnan PT. Apolipoprotein e genotypes are associated with lipid-lowering responses to statin treatment in diabetes: A go-darts study. *Pharmacogenet Genomics*. 2008;18:279-287

257. Donnelly LA, Doney AS, Tavendale R, Lang CC, Pearson ER, Colhoun HM, McCarthy MI, Hattersley AT, Morris AD, Palmer CN. Common nonsynonymous substitutions in slco1b1 predispose to statin intolerance in routinely treated individuals with type 2 diabetes: A go-darts study. *Clinical pharmacology and therapeutics*. 2011;89:210-216

258. Evans JM, Doney AS, AlZadjali MA, Ogston SA, Petrie JR, Morris AD, Struthers AD, Wong AK, Lang CC. Effect of metformin on mortality in patients with heart failure and type 2 diabetes mellitus. *Am J Cardiol*. 2010;106:1006-1010

259. Doney AS, Lee S, Leese GP, Morris AD, Palmer CN. Increased cardiovascular morbidity and mortality in type 2 diabetes is associated with the glutathione s transferase theta-null genotype: A go-darts study. *Circulation*. 2005;111:2927-2934

260. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, Hindy G, Holm H, Ding EL, Johnson T, Schunkert H, Samani NJ, Clarke R, Hopewell JC, Thompson JF, Li M, Thorleifsson G, Newton-Cheh C, Musunuru K, Pirruccello JP, Saleheen D, Chen L, Stewart AF, Schillert A, Thorsteinsdottir U, Thorgeirsson G, Anand S, Engert JC, Morgan T, Spertus J, Stoll M, Berger K, Martinelli N, Girelli D, McKeown PP, Patterson CC, Epstein SE, Devaney J, Burnett MS, Mooser V, Ripatti S, Surakka I, Nieminen MS, Sinisalo J, Lokki ML, Perola M, Havulinna A, de Faire U, Gigante B, Ingelsson E, Zeller T, Wild P, de Bakker PI, Klungel OH, Maitland-van der Zee AH, Peters BJ, de Boer A, Grobbee DE, Kamphuisen PW, Deneer VH, Elbers CC, Onland-Moret NC, Hofker MH, Wijmenga C, Verschuren WM, Boer JM, van der Schouw YT, Rasheed A, Frossard P, Demissie S, Willer C, Do R, Ordovas JM, Abecasis GR, Boehnke M, Mohlke KL, Daly MJ, Guiducci C, Burtt NP, Surti A, Gonzalez E, Purcell S, Gabriel S, Marrugat J, Peden J, Erdmann J, Diemert P, Willenborg C, Konig IR, Fischer M, Hengstenberg C, Ziegler A, Buysschaert I, Lambrechts D, Van de Werf F, Fox KA, El Mokhtari NE, Rubin D, Schrezenmeir J, Schreiber S, Schafer A, Danesh J, Blankenberg S, Roberts R, McPherson R, Watkins H, Hall AS, Overvad K, Rimm E, Boerwinkle E, Tybjaerg-Hansen A, Cupples LA, Reilly MP, Melander O, Mannucci PM, Ardissino D, Siscovick D, Elosua R, Stefansson K, O'Donnell CJ, Salomaa V, Rader DJ, Peltonen L, Schwartz SM, Altshuler D, Kathiresan S. Plasma hdl cholesterol and risk of myocardial infarction: A mendelian randomisation study. *Lancet*. 2012

261. Flynn RW, Macdonald TM, Schembri N, Murray GD, Doney AS. Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes. *Pharmacoepidemiology and drug safety*. 2010;19:843-847

262. Kullo IJ, Fan J, Pathak J, Savova GK, Ali Z, Chute CG. Leveraging informatics for genetic studies: Use of the electronic medical record to enable a genome-wide association study of peripheral arterial disease. *Journal of the American Medical Informatics Association : JAMIA*. 2010;17:568-574

263. Peach G, Griffin M, Jones K, Thompson M, Hinchliffe R. Diagnosis and management of peripheral arterial disease. *BMJ (Clinical research ed.)*. 2012;345

264. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011;1:457-470

265. Middelberg RP, Ferreira MA, Henders AK, Heath AC, Madden PA, Montgomery GW, Martin NG, Whitfield JB. Genetic variants in lpl, oasl and tomm40/apoe-c1-c2-c4 genes are associated with multiple cardiovascular-related traits. *BMC medical genetics*. 2011;12:123

266. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012

267. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Magi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparso T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proenca C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Bottcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllensten U, Hadjadj S, Hallmans G, Hammond N, Han X, Hartikainen AL, Hassanali N, Hayward C, Heath SC,

Hercberg S, Herder C, Hicks AA, Hillman DR, Hingorani AD, Hofman A, Hui J, Hung J, Isomaa B, Johnson PR, Jorgensen T, Jula A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoeur C, Li Y, Lyssenko V, Mahley R, Mangino M, Manning AK, Martinez-Larrad MT, McAteer JB, McCulloch LJ, McPherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Morken MA, Mukherjee S, Naitza S, Narisu N, Neville MJ, Oostra BA, Orru M, Pakyz R, Palmer CN, Paolisso G, Pattaro C, Pearson D, Peden JF, Pedersen NL, Perola M, Pfeiffer AF, Pichler I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rathmann W, Rayner NW, Rice K, Ripatti S, Rivadeneira F, Roden M, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Scott LJ, Seedorf U, Sharp SJ, Shields B, Sigurethsson G, Sijbrands EJ, Silveira A, Simpson L, Singleton A, Smith NL, Sovio U, Swift A, Syddall H, Syvanen AC, Tanaka T, Thorand B, Tichet J, Tonjes A, Tuomi T, Uitterlinden AG, van Dijk KW, van Hoek M, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N, Waeber G, Wagner PJ, Walley A, Walters GB, Ward KL, Watkins H, Weedon MN, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zeggini E, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC, Borecki IB, Loos RJ, Meneton P, Magnusson PK, Nathan DM, Williams GH, Hattersley AT, Silander K, Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C, Dedoussis GV, Serrano-Rios M, Morris AD, Lind L, Palmer LJ, Hu FB, Franks PW, Ebrahim S, Marmot M, Kao WH, Pankow JS, Sampson MJ, Kuusisto J, Laakso M, Hansen T, Pedersen O, Pramstaller PP, Wichmann HE, Illig T, Rudan I, Wright AF, Stumvoll M, Campbell H, Wilson JF, Bergman RN, Buchanan TA, Collins FS, Mohlke KL, Tuomilehto J, Valle TT, Altshuler D, Rotter JI, Siscovick DS, Penninx BW, Boomsma DI, Deloukas P, Spector TD, Frayling TM, Ferrucci L, Kong A, Thorsteinsdottir U, Stefansson K, van Duijn CM, Aulchenko YS, Cao A, Scuteri A, Schlessinger D, Uda M, Ruokonen A, Jarvelin MR, Waterworth DM, Vollenweider P, Peltonen L, Mooser V, Abecasis GR, Wareham NJ, Sladek R, Froguel P, Watanabe RM, Meigs JB, Groop L, Boehnke M, McCarthy MI, Florez JC, Barroso I. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42:105-116

268. Cluett C, McDermott MM, Guralnik J, Ferrucci L, Bandinelli S, Miljkovic I, Zmuda JM, Li R, Tranah G, Harris T, Rice N, Henley W, Frayling TM, Murray A, Melzer D. The 9p21 myocardial infarction risk allele increases risk of peripheral artery disease in older people. *Circ Cardiovasc Genet*. 2009;2:347-353

269. Talmud PJ, Cooper JA, Palmen J, Lovering R, Drenos F, Hingorani AD, Humphries SE. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of chd in healthy middle-aged men. *Clin Chem*. 2008;54:467-474

270. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488-1491

271. Chen Z, Qian Q, Ma G, Wang J, Zhang X, Feng Y, Shen C, Yao Y. A common variant on chromosome 9p21 affects the risk of early-onset coronary artery disease. *Mol Biol Rep*. 2009;36:889-893

272. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H,

Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-1493

273. Helgadottir A, Thorleifsson G, Magnusson KP, Gretarsdottir S, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJE, Blankensteijn JD, Ronkainen A, Jaaskelainen JE, Kyo Y, Lenk GM, Sakalihasan N, Kostulas K, Gottsater A, Flex A, Stefansson H, Hansen T, Andersen G, Weinsheimer S, Borch-Johnsen K, Jorgensen T, Shah SH, Quyyumi AA, Granger CB, Reilly MP, Austin H, Levey AI, Vaccarino V, Palsdottir E, Walters GB, Jonsdottir T, Snorradottir S, Magnusdottir D, Gudmundsson G, Ferrell RE, Sveinbjornsdottir S, Hernesniemi J, Niemela M, Limet R, Andersen K, Sigurdsson G, Benediktsson R, Verhoeven ELG, Teijink JAW, Grobbee DE, Rader DJ, Collier DA, Pedersen O, Pola R, Hillert J, Lindblad B, Valdimarsson EM, Magnadottir HB, Wijmenga C, Tromp G, Baas AF, Ruigrok YM, van Rij AM, Kuivaniemi H, Powell JT, Matthiasson SE, Gulcher JR, Thorgeirsson G, Kong A, Thorsteinsdottir U, Stefansson K. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet*. 2008;40:217-224

274. Assimes TL, Knowles JW, Basu A, Iribarren C, Southwick A, Tang H, Absher D, Li J, Fair JM, Rubin GD, Sidney S, Fortmann SP, Go AS, Hlatky MA, Myers RM, Risch N, Quertermous T. Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic advance study. *Hum. Mol. Genet.* 2008;17:2320-2328

275. Matarin M, Brown WM, Singleton A, Hardy JA, Meschia JF. Whole genome analyses suggest ischemic stroke and heart disease share an association with polymorphisms on chromosome 9p21. *Stroke*. 2008;39:1586-1589

276. Smith S, Greenland P, Grundy S. Aha conference proceedings. Prevention conference v: Beyond secondary prevention: Identifying the high-risk patient for primary prevention: Executive summary. American heart association. *Circulation*. 2000;101:111 - 116

277. Catalano M, Cortelazzo A, Yilmaz Y, Perilli E, Carzaniga G, Emanuele E. The lpa gene c93t polymorphism influences plasma lipoprotein(a) levels and is independently associated with susceptibility to peripheral arterial disease. *Clin Chim Acta*. 2008;387:109-112

278. Smolders B, Lemmens R, Thijs V. Lipoprotein (a) and stroke: A meta-analysis of observational studies. *Stroke*. 2007;38:1959-1966

279. Vaxillaire M, Cavalcanti-Proenca C, Dechaume A, Tichet J, Marre M, Balkau B, Froguel P. The common p446l polymorphism in gckr inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the desir prospective general french population. *Diabetes*. 2008;57:2253-2257

280. Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, Lyssenko V, Bouatia-Naji N, Dupuis J, Jackson AU, Kao WH, Li M, Glazer NL, Manning AK, Luan J, Stringham HM, Prokopenko I, Johnson T, Grarup N, Boesgaard TW, Lecoeur C, Shrader P, O'Connell J, Ingelsson E, Couper DJ, Rice K, Song K, Andreasen CH, Dina C, Kottgen A, Le Bacquer O, Pattou F, Taneera J, Steinthorsdottir V, Rybin D, Ardlie K, Sampson M, Qi L, van Hoek M, Weedon MN, Aulchenko YS, Voight BF, Grallert H, Balkau B, Bergman RN, Bielinski SJ, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Bottcher Y, Brunner E, Buchanan TA,

Bumpstead SJ, Cavalcanti-Proenca C, Charpentier G, Chen YD, Chines PS, Collins FS, Cornelis M, G JC, Delplanque J, Doney A, Egan JM, Erdos MR, Firmann M, Forouhi NG, Fox CS, Goodarzi MO, Graessler J, Hingorani A, Isomaa B, Jorgensen T, Kivimaki M, Kovacs P, Krohn K, Kumari M, Lauritzen T, Levy-Marchal C, Mayor V, McAteer JB, Meyre D, Mitchell BD, Mohlke KL, Morken MA, Narisu N, Palmer CN, Pakyz R, Pascoe L, Payne F, Pearson D, Rathmann W, Sandbaek A, Sayer AA, Scott LJ, Sharp SJ, Sijbrands E, Singleton A, Siscovick DS, Smith NL, Sparso T, Swift AJ, Syddall H, Thorleifsson G, Tonjes A, Tuomi T, Tuomilehto J, Valle TT, Waeber G, Walley A, Waterworth DM, Zeggini E, Zhao JH, Illig T, Wichmann HE, Wilson JF, van Duijn C, Hu FB, Morris AD, Frayling TM, Hattersley AT, Thorsteinsdottir U, Stefansson K, Nilsson P, Syvanen AC, Shuldiner AR, Walker M, Bornstein SR, Schwarz P, Williams GH, Nathan DM, Kuusisto J, Laakso M, Cooper C, Marmot M, Ferrucci L, Mooser V, Stumvoll M, Loos RJ, Altshuler D, Psaty BM, Rotter JI, Boerwinkle E, Hansen T, Pedersen O, Florez JC, McCarthy MI, Boehnke M, Barroso I, Sladek R, Froguel P, Meigs JB, Groop L, Wareham NJ, Watanabe RM. Genetic variation in gipr influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet*. 2010;42:142-148

281. Sparso T, Andersen G, Nielsen T, Burgdorf KS, Gjesing AP, Nielsen AL, Albrechtsen A, Rasmussen SS, Jorgensen T, Borch-Johnsen K, Sandbaek A, Lauritzen T, Madsbad S, Hansen T, Pedersen O. The gckr rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and ogtt-related insulinaemia, and reduced risk of type 2 diabetes. *Diabetologia*. 2008;51:70-75

282. Palmer CN, Maglio C, Pirazzi C, Burza MA, Adiels M, Burch L, Donnelly LA, Colhoun H, Doney AS, Dillon JF, Pearson ER, McCarthy M, Hattersley AT, Frayling T, Morris AD, Peltonen M, Svensson PA, Jacobson P, Boren J, Sjostrom L, Carlsson LM, Romeo S. Paradoxical lower serum triglyceride levels and higher type 2 diabetes mellitus susceptibility in obese individuals with the pnpla3 148m variant. *PLoS One*. 2012;7:e39362

283. Sullivan FM, Donnan PT, Love T. Whole population secondary prevention of coronary heart disease in scotland: The hearts database. *Studies in health technology and informatics*. 2004;107:1227-1229

284. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837-1847

285. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, Konig IR, Cazier JB, Johansson A, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikainen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney AS, Mokhtari NE, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Muller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schafer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, Schoot CE, Wagner PJ, Wells GA, Wild PS, Yang TP, Amouyel P,

Arveiler D, Basart H, Boehnke M, Boerwinkle E, Brambilla P, Cambien F, Cupples AL, de Faire U, Dehghan A, Diemert P, Epstein SE, Evans A, Ferrario MM, Ferrieres J, Gauguier D, Go AS, Goodall AH, Gudnason V, Hazen SL, Holm H, Iribarren C, Jang Y, Kahonen M, Kee F, Kim HS, Klopp N, Koenig W, Kratzer W, Kuulasmaa K, Laakso M, Laaksonen R, Lind L, Ouwehand WH, Parish S, Park JE, Pedersen NL, Peters A, Quertermous T, Rader DJ, Salomaa V, Schadt E, Shah SH, Sinisalo J, Stark K, Stefansson K, Tregouet DA, Virtamo J, Wallentin L, Wareham N, Zimmermann ME, Nieminen MS, Hengstenberg C, Sandhu MS, Pastinen T, Syvanen AC, Hovingh GK, Dedoussis G, Franks PW, Lehtimaki T, Metspalu A, Zalloua PA, Siegbahn A, Schreiber S, Ripatti S, Blankenberg SS, Perola M, Clarke R, Boehm BO, O'Donnell C, Reilly MP, Marz W, Collins R, Kathiresan S, Hamsten A, Kooner JS, Thorsteinsdottir U, Danesh J, Palmer CN, Roberts R, Watkins H, Schunkert H, Samani NJ. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2012;45:25-33

286. Leusink M, Onland-Moret NC, Asselbergs FW, Ding B, Kotti S, Zuydam NRv, Papp AC, Danchin N, Donnelly L, Morris AD, Chasman DI, Doevendans PAFM, Klungel OH, Ridker PM, Gilst WHv, Simon T, Nyberg F, Palmer CNA, Sadee W, Harst Pvd, Bakker PIWd, Boer Ad, Verstuyft C, Zee AHM-vd. Cholesteryl ester transfer protein (cetp) polymorphisms, statin use, and their impact on cholesterol levels and cardiovascular events. 2013

287. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44:981-990

288. Dupont WD, Plummer WD, Jr. Power and sample size calculations. A review and computer program. *Controlled clinical trials*. 1990;11:116-128

289. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *NATURE*. 2011;472:57-63

290. Vaara S, Nieminen MS, Lokki ML, Perola M, Pussinen PJ, Allonen J, Parkkonen O, Sinisalo J. Cohort profile: The corogene study. *Int J Epidemiol*. 2012;41:1265-1271

291. Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S, Piskackova T, Balascak I, Peltonen L, Jakkula E, Rehnstrom K, Lathrop M, Heath S, Galan P, Schreiber S, Meitinger T, Pfeufer A, Wichmann HE, Melegh B, Polgar N, Toniolo D, Gasparini P, D'Adamo P, Klovins J, Nikitina-Zake L, Kucinskas V, Kasnauskiene J, Lubinski J, Debniak T, Limborska S, Khrunin A, Estivill X, Rabionet R, Marsal S, Julia A, Antonarakis SE, Deutsch S, Borel C, Attar H, Gagnebin M, Macek M, Krawczak M, Remm M, Metspalu A. Genetic structure of europeans: A view from the north-east. *PLoS One*. 2009;4:e5472

292. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, West K, Kashuk C, Akyol M, Perz S, Jalilzadeh S, Illig T, Gieger C, Guo CY, Larson MG, Wichmann HE, Marban E, O'Donnell CJ, Hirschhorn JN, Kaab S, Spooner PM, Meitinger T, Chakravarti A. A common genetic variant in the nos1 regulator nos1ap modulates cardiac repolarization. *Nat Genet*. 2006;38:644-651

293. Balbarini A, Buttitta F, Limbruno U, Petronio AS, Baglini R, Strata G, Mariotti R, Ciccone M, Mariani M. Usefulness of carotid intima-media thickness measurement and peripheral b-mode ultrasound scan in the clinical screening of patients with coronary artery disease. *Angiology*. 2000;51:269-279

294. Samnegard A, Silveira A, Lundman P, Boquist S, Odeberg J, Hulthe J, McPheat W, Tornvall P, Bergstrand L, Ericsson CG, Hamsten A, Eriksson P. Serum matrix metalloproteinase-3 concentration is influenced by mmp-3 -1612 5a/6a promoter genotype and associated with myocardial infarction. *J Intern Med*. 2005;258:411-419

295. Evans A, Salomaa V, Kulathinal S, Asplund K, Cambien F, Ferrario M, Perola M, Peltonen L, Shields D, Tunstall-Pedoe H, Kuulasmaa K. Morgam (an international pooling of cardiovascular cohorts). *Int J Epidemiol*. 2005;34:21-27

296. Theodoraki EV, Nikopensius T, Suhorutsenko J, Peppes V, Fili P, Kolovou G, Papamikos V, Richter D, Zakopoulos N, Krjutskov K, Metspalu A, Dedoussis GV. Fibrinogen beta variants confer protection against coronary artery disease in a greek case-control study. *BMC medical genetics*. 2010;11:28

297. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, Hamalainen E, Jousilahti P, Kangas AJ, Mannisto S, Savolainen MJ, Jula A, Leiviska J, Palotie A, Salomaa V, Perola M, Ala-Korpela M, Peltonen L. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Molecular systems biology*. 2010;6:441

298. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*. 2007;316:1341-1345

299. Wang L, Zheng J, Bai X, Liu B, Liu CJ, Xu Q, Zhu Y, Wang N, Kong W, Wang X. Adamts-7 mediates vascular smooth muscle cell migration and neointima formation in balloon-injured rat arteries. *Circ Res*. 2009;104:688-698

300. Hughes SE. Differential expression of the fibroblast growth factor receptor (fgfr) multigene family in normal human adult tissues. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. 1997;45:1005-1019

301. Jackson CL, Reidy MA. Basic fibroblast growth factor: Its role in the control of smooth muscle cell migration. *The American journal of pathology*. 1993;143:1024-1031

302. Sabatel C, Cornet AM, Tabruyn SP, Malvaux L, Castermans K, Martial JA, Struman I. Sprouty1, a new target of the angiostatic agent 16k prolactin, negatively regulates angiogenesis. *Mol Cancer*. 2010;9:231

303. Martin JE, Alizadeh BZ, Gonzalez-Gay MA, Balsa A, Pascual-Salcedo D, Fernandez-Gutierrez B, Raya E, Franke L, van't Slot R, Coenen MJ, van Riel P, Radstake TR, Koeleman BP, Martin J. Identification of the oxidative stress-related gene msra as a rheumatoid arthritis susceptibility locus by genome-wide pathway analysis. *Arthritis and rheumatism*. 2010;62:3183-3190

304. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C, Lettre G, Lim N, Lyon HN, McCarroll SA, Papadakis K, Qi L, Randall JC, Roccasecca RM, Sanna S, Scheet P, Weedon MN, Wheeler E, Zhao JH, Jacobs LC, Prokopenko I, Soranzo N, Tanaka T, Timpson NJ, Almgren P, Bennett A, Bergman RN, Bingham SA, Bonnycastle LL, Brown M, Burtt NP, Chines P, Coin L, Collins FS, Connell JM, Cooper C, Smith GD, Dennison EM, Deodhar P, Elliott P, Erdos MR, Estrada K, Evans DM, Gianniny L, Gieger C, Gillson CJ, Guiducci C, Hackett R, Hadley D, Hall AS, Havulinna AS, Hebebrand J, Hofman A, Isomaa B, Jacobs KB, Johnson T, Jousilahti P, Jovanovic Z, Khaw KT, Kraft P, Kuokkanen M, Kuusisto J, Laitinen J, Lakatta EG, Luan J, Luben RN, Mangino M, McArdle WL, Meitinger T, Mulas A, Munroe PB, Narisu N, Ness AR, Northstone K, O'Rahilly S, Purmann C, Rees MG, Ridderstrale M, Ring SM, Rivadeneira F, Ruokonen A, Sandhu MS, Saramies J, Scott LJ, Scuteri A, Silander K, Sims MA, Song K, Stephens J, Stevens S, Stringham HM, Tung YC, Valle TT, Van Duijn CM, Vimaleswaran KS, Vollenweider P, Waeber G, Wallace C, Watanabe RM, Waterworth DM, Watkins N, Witteman JC, Zeggini E, Zhai G, Zillikens MC, Altshuler D, Caulfield MJ, Chanock SJ, Farooqi IS, Ferrucci L, Guralnik JM, Hattersley AT, Hu FB, Jarvelin MR, Laakso M, Mooser V, Ong KK, Ouwehand WH, Salomaa V, Samani NJ, Spector TD, Tuomi T, Tuomilehto J, Uda M, Uitterlinden AG, Wareham NJ, Deloukas P, Frayling TM, Groop LC, Hayes RB, Hunter DJ, Mohlke KL, Peltonen L, Schlessinger D, Strachan DP, Wichmann HE, McCarthy MI, Boehnke M, Barroso I, Abecasis GR, Hirschhorn JN. Six new loci associated

with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet*. 2009;41:25-34

305. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Magi R, Randall JC, Vedantam S, Winkler TW, Qi L, Workalemahu T, Heid IM, Steinthorsdottir V, Stringham HM, Weedon MN, Wheeler E, Wood AR, Ferreira T, Weyant RJ, Segre AV, Estrada K, Liang L, Nemesh J, Park JH, Gustafsson S, Kilpelainen TO, Yang J, Bouatia-Naji N, Esko T, Feitosa MF, Kutalik Z, Mangino M, Raychaudhuri S, Scherag A, Smith AV, Welch R, Zhao JH, Aben KK, Absher DM, Amin N, Dixon AL, Fisher E, Glazer NL, Goddard ME, Heard-Costa NL, Hoesel V, Hottenga JJ, Johansson A, Johnson T, Ketkar S, Lamina C, Li S, Moffatt MF, Myers RH, Narisu N, Perry JR, Peters MJ, Preuss M, Ripatti S, Rivadeneira F, Sandholt C, Scott LJ, Timpson NJ, Tyrer JP, van Wingerden S, Watanabe RM, White CC, Wiklund F, Barlassina C, Chasman DI, Cooper MN, Jansson JO, Lawrence RW, Pellikka N, Prokopenko I, Shi J, Thiering E, Alavere H, Alibrandi MT, Almgren P, Arnold AM, Aspelund T, Atwood LD, Balkau B, Balmforth AJ, Bennett AJ, Ben-Shlomo Y, Bergman RN, Bergmann S, Biebermann H, Blakemore AI, Boes T, Bonnycastle LL, Bornstein SR, Brown MJ, Buchanan TA, Busonero F, Campbell H, Cappuccio FP, Cavalcanti-Proenca C, Chen YD, Chen CM, Chines PS, Clarke R, Coin L, Connell J, Day IN, den Heijer M, Duan J, Ebrahim S, Elliott P, Elosua R, Eiriksdottir G, Erdos MR, Eriksson JG, Facheris MF, Felix SB, Fischer-Posovszky P, Folsom AR, Friedrich N, Freimer NB, Fu M, Gaget S, Gejman PV, Geus EJ, Gieger C, Gjesing AP, Goel A, Goyette P, Grallert H, Grassler J, Greenawalt DM, Groves CJ, Gudnason V, Guiducci C, Hartikainen AL, Hassanali N, Hall AS, Havulinna AS, Hayward C, Heath AC, Hengstenberg C, Hicks AA, Hinney A, Hofman A, Homuth G, Hui J, Igl W, Iribarren C, Isomaa B, Jacobs KB, Jarick I, Jewell E, John U, Jorgensen T, Jousilahti P, Jula A, Kaakinen M, Kajantie E, Kaplan LM, Kathiresan S, Kettunen J, Kinnunen L, Knowles JW, Kolcic I, Konig IR, Koskinen S, Kovacs P, Kuusisto J, Kraft P, Kvaloy K, Laitinen J, Lantieri O, Lanzani C, Launer LJ, Lecoeur C, Lehtimaki T, Lettre G, Liu J, Lokki ML, Lorentzon M, Luben RN, Ludwig B, Manunta P, Marek D, Marre M, Martin NG, McArdle WL, McCarthy A, McKnight B, Meitinger T, Melander O, Meyre D, Midthjell K, Montgomery GW, Morken MA, Morris AP, Mulic R, Ngwa JS, Nelis M, Neville MJ, Nyholt DR, O'Donnell CJ, O'Rahilly S, Ong KK, Oostra B, Pare G, Parker AN, Perola M, Pichler I, Pietilainen KH, Platou CG, Polasek O, Pouta A, Rafelt S, Raitakari O, Rayner NW, Ridderstrale M, Rief W, Ruokonen A, Robertson NR, Rzehak P, Salomaa V, Sanders AR, Sandhu MS, Sanna S, Saramies J, Savolainen MJ, Scherag S, Schipf S, Schreiber S, Schunkert H, Silander K, Sinisalo J, Siscovick DS, Smit JH, Soranzo N, Sovio U, Stephens J, Surakka I, Swift AJ, Tammesoo ML, Tardif JC, Teder-Laving M, Teslovich TM, Thompson JR, Thomson B, Tonjes A, Tuomi T, van Meurs JB, van Ommen GJ, Vatin V, Viikari J, Visvikis-Siest S, Vitart V, Vogel CI, Voight BF, Waite LL, Wallaschofski H, Walters GB, Widen E, Wiegand S, Wild SH, Willemsen G, Witte DR, Witteman JC, Xu J, Zhang Q, Zgaga L, Ziegler A, Zitting P, Beilby JP, Farooqi IS, Hebebrand J, Huikuri HV, James AL, Kahonen M, Levinson DF, Macciardi F, Nieminen MS, Ohlsson C, Palmer LJ, Ridker PM, Stumvoll M, Beckmann JS, Boeing H, Boerwinkle E, Boomsma DI, Caulfield MJ, Chanock SJ, Collins FS, Cupples LA, Smith GD, Erdmann J, Froguel P, Gronberg H, Gyllensten U, Hall P, Hansen T, Harris TB, Hattersley AT, Hayes RB, Heinrich J, Hu FB, Hveem K, Illig T, Jarvelin MR, Kaprio J, Karpe F, Khaw KT, Kiemeney LA, Krude H, Laakso M, Lawlor DA, Metspalu A, Munroe PB, Ouwehand WH, Pedersen O, Penninx BW, Peters A, Pramstaller PP, Quertermous T, Reinehr T, Rissanen A, Rudan I, Samani NJ, Schwarz PE, Shuldiner AR, Spector TD, Tuomilehto J, Uda M, Uitterlinden A, Valle TT, Wabitsch M, Waeber G, Wareham NJ, Watkins H, Wilson JF, Wright

AF, Zillikens MC, Chatterjee N, McCarroll SA, Purcell S, Schadt EE, Visscher PM, Assimes TL, Borecki IB, Deloukas P, Fox CS, Groop LC, Haritunians T, Hunter DJ, Kaplan RC, Mohlke KL, O'Connell JR, Peltonen L, Schlessinger D, Strachan DP, van Duijn CM, Wichmann HE, Frayling TM, Thorsteinsdottir U, Abecasis GR, Barroso I, Boehnke M, Stefansson K, North KE, McCarthy MI, Hirschhorn JN, Ingelsson E, Loos RJ. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42:937-948

306. Rull A, Garcia R, Fernandez-Sender L, Garcia-Heredia A, Aragones G, Beltran-Debon R, Marsillach J, Alegret JM, Martin-Paredero V, Mackness B, Mackness M, Joven J, Camps J. Serum paraoxonase-3 concentration is associated with insulin sensitivity in peripheral artery disease and with inflammation in coronary artery disease. *Atherosclerosis*. 2012;220:545-551

307. Rajkovic MG, Rumora L, Barisic K. The paraoxonase 1, 2 and 3 in humans. *Biochemia medica : casopis Hrvatskoga drustva medicinskih biokemicara / HDMB*. 2011;21:122-130

308. Fuhrman B, Gantman A, Aviram M. Paraoxonase 1 (pon1) deficiency in mice is associated with reduced expression of macrophage sr-bi and consequently the loss of hdl cytoprotection against apoptosis. *Atherosclerosis*. 2010;211:61-68

309. Fuhrman B. Regulation of hepatic paraoxonase-1 expression. *Journal of lipids*. 2012;2012:684010

310. She ZG, Chen HZ, Yan Y, Li H, Liu DP. The human paraoxonase gene cluster as a target in the treatment of atherosclerosis. *Antioxidants & redox signaling*. 2012;16:597-632

311. Fujimoto Y, Itabe H, Sakai J, Makita M, Noda J, Mori M, Higashi Y, Kojima S, Takano T. Identification of major proteins in the lipid droplet-enriched fraction isolated from the human hepatocyte cell line huh7. *Biochim Biophys Acta*. 2004;1644:47-59

312. Fujimoto Y, Itabe H, Kinoshita T, Homma KJ, Onoduka J, Mori M, Yamaguchi S, Makita M, Higashi Y, Yamashita A, Takano T. Involvement of acsl in local synthesis of neutral lipids in cytoplasmic lipid droplets in human hepatocyte huh7. *J Lipid Res*. 2007;48:1280-1292

313. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JS, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, Green RC, Rogaeva E, St George-Hyslop P, Arnold SE, Barber R, Beach T, Bigio EH, Bowen JD, Boxer A, Burke JR, Cairns NJ, Carlson CS, Carney RM, Carroll SL, Chui HC, Clark DG, Corneveaux J, Cotman CW, Cummings JL, DeCarli C, DeKosky ST, Diaz-Arrastia R, Dick M, Dickson DW, Ellis WG, Faber KM, Fallon KB, Farlow MR, Ferris S, Frosch MP, Galasko DR, Ganguli M, Gearing M, Geschwind DH, Ghetti B, Gilbert JR, Gilman S, Giordani B, Glass JD, Growdon JH, Hamilton RL, Harrell LE, Head E, Honig LS, Hulette CM, Hyman BT, Jicha GA, Jin LW, Johnson N, Karlawish J, Karydas A, Kaye JA, Kim R, Koo EH, Kowall NW, Lah JJ, Levey AI, Lieberman AP, Lopez OL, Mack WJ, Marson DC, Martiniuk F, Mash DC, Masliah E, McCormick WC, McCurry SM, McDavid AN, McKee AC, Mesulam M, Miller BL, Miller CA, Miller JW, Parisi JE, Perl DP, Peskind E, Petersen RC, Poon WW, Quinn JF, Rajbhandary RA, Raskind M, Reisberg B, Ringman JM, Roberson ED, Rosenberg RN, Sano M, Schneider LS, Seeley W,

Shelanski ML, Slifer MA, Smith CD, Sonnen JA, Spina S, Stern RA, Tanzi RE, Trojanowski JQ, Troncoso JC, Van Deerlin VM, Vinters HV, Vonsattel JP, Weintraub S, Welsh-Bohmer KA, Williamson J, Woltjer RL, Cantwell LB, Dombroski BA, Beekly D, Lunetta KL, Martin ER, Kamboh MI, Saykin AJ, Reiman EM, Bennett DA, Morris JC, Montine TJ, Goate AM, Blacker D, Tsuang DW, Hakonarson H, Kukull WA, Foroud TM, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD. Common variants at ms4a4/ms4a6e, cd2ap, cd33 and epha1 are associated with late-onset alzheimer's disease. *Nat Genet*. 2011;43:436-441

314. Abbott GW, Sesti F, Splawski I, Buck ME, Lehmann MH, Timothy KW, Keating MT, Goldstein SA. Mirp1 forms ikr potassium channels with herg and is associated with cardiac arrhythmia. *Cell*. 1999;97:175-187

315. Koshelev M, Sarma S, Price RE, Wehrens XH, Cooper TA. Heart-specific overexpression of cugbp1 reproduces functional and molecular abnormalities of myotonic dystrophy type 1. *Hum Mol Genet*. 2010;19:1066-1075

316. Luo X, Hojayev B, Jiang N, Wang ZV, Tandan S, Rakalin A, Rothermel BA, Gillette TG, Hill JA. Stim1-dependent store-operated ca(2)(+) entry is required for pathological cardiac hypertrophy. *Journal of molecular and cellular cardiology*. 2012;52:136-147

317. Massague J, Wotton D. Transcriptional control by the tgf-beta/smad signaling system. *Embo J*. 2000;19:1745-1754

318. Kalinina N, Agrotis A, Antropova Y, Ilyinskaya O, Smirnov V, Tararak E, Bobik A. Smad expression in human atherosclerotic lesions: Evidence for impaired tgf-beta/smad signaling in smooth muscle cells of fibrofatty lesions. *Arterioscler Thromb Vasc Biol*. 2004;24:1391-1396

319. Li JH, Huang XR, Zhu HJ, Oldfield M, Cooper M, Truong LD, Johnson RJ, Lan HY. Advanced glycation end products activate smad signaling via tgf-beta-dependent and independent mechanisms: Implications for diabetic renal and vascular disease. *FASEB J*. 2004;18:176-178

320. Hauner H. The mode of action of thiazolidinediones. *Diabetes Metab Res Rev*. 2002;18 Suppl 2:S10-15

321. Lincoff AM, Wolski K, Nicholls SJ, Nissen SE. Pioglitazone and risk of cardiovascular events in patients with type 2 diabetes mellitus: A meta-analysis of randomized trials. *JAMA*. 2007;298:1180-1188

322. Nissen SE, Nicholls SJ, Wolski K, Nesto R, Kupfer S, Perez A, Jure H, De Larochelliere R, Staniloae CS, Mavromatis K, Saw J, Hu B, Lincoff AM, Tuzcu EM. Comparison of pioglitazone vs glimepiride on progression of coronary atherosclerosis in patients with type 2 diabetes: The periscope randomized controlled trial. *JAMA*. 2008;299:1561-1573

323. Benigni A, Zoja C, Tomasoni S, Campana M, Corna D, Zanchi C, Gagliardini E, Garofano E, Rottoli D, Ito T, Remuzzi G. Transcriptional regulation of nephrin gene by peroxisome proliferator-activated receptor-gamma agonist: Molecular mechanism of the antiproteinuric effect of pioglitazone. *J Am Soc Nephrol*. 2006;17:1624-1632

324. Kanjanabuch T, Ma LJ, Chen J, Pozzi A, Guan Y, Mundel P, Fogo AB. Ppar-gamma agonist protects podocytes from injury. *Kidney Int*. 2007;71:1232-1239

325. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzosi MG, Tognoni G, Seedorf U, Rust S, Eriksson P, Hamsten A, Farrall M, Watkins H. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked snps in the anril locus on chromosome 9p. *Hum Mol Genet*. 2008;17:806-814

326. Cigolini M, Iagulli MP, Miconi V, Galiotto M, Lombardi S, Targher G. Serum 25-hydroxyvitamin d3 concentrations and prevalence of cardiovascular disease among type 2 diabetic patients. *Diabetes Care*. 2006;29:722-724

327. Hypponen E, Power C. Vitamin d status and glucose homeostasis in the 1958 british birth cohort: The role of obesity. *Diabetes Care*. 2006;29:2244-2246

328. Scragg R, Sowers M, Bell C. Serum 25-hydroxyvitamin d, ethnicity, and blood pressure in the third national health and nutrition examination survey. *Am J Hypertens*. 2007;20:713-719

329. Scragg R, Holdaway I, Singh V, Metcalf P, Baker J, Dryson E. Serum 25-hydroxyvitamin d3 levels decreased in impaired glucose tolerance and diabetes mellitus. *Diabetes Res Clin Pract*. 1995;27:181-188

330. Wang TJ, Pencina MJ, Booth SL, Jacques PF, Ingelsson E, Lanier K, Benjamin EJ, D'Agostino RB, Wolf M, Vasan RS. Vitamin d deficiency and risk of cardiovascular disease. *Circulation*. 2008;117:503-511

331. Dobnig H, Pilz S, Scharnagl H, Renner W, Seelhorst U, Wellnitz B, Kinkeldei J, Boehm BO, Weihrauch G, Maerz W. Independent association of low serum 25-hydroxyvitamin d and 1,25-dihydroxyvitamin d levels with all-cause and cardiovascular mortality. *Arch Intern Med*. 2008;168:1340-1349

332. Davey Smith G, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1-22

333. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med*. 2008;5:e177

334. Wensley F, Gao P, Burgess S, Kaptoge S, Di Angelantonio E, Shah T, Engert JC, Clarke R, Davey-Smith G, Nordestgaard BG, Saleheen D, Samani NJ, Sandhu M, Anand S, Pepys MB, Smeeth L, Whittaker J, Casas JP, Thompson SG, Hingorani AD, Danesh J. Association between c reactive protein and coronary heart disease: Mendelian randomisation analysis based on individual participant data. *BMJ*. 2011;342:d548

335. Sarwar N, Sandhu MS, Ricketts SL, Butterworth AS, Di Angelantonio E, Boekholdt SM, Ouwehand W, Watkins H, Samani NJ, Saleheen D, Lawlor D, Reilly MP, Hingorani AD, Talmud PJ, Danesh J. Triglyceride-mediated pathways and coronary disease: Collaborative analysis of 101 studies. *Lancet*. 2010;375:1634-1639

336. Schleinitz D, Distefano JK, Kovacs P. Targeted snp genotyping using the taqman(r) assay. *Methods mol biol*. 2011:77-87.

337. Therneau T, Lumley T. Survival: Survival analysis, including penalised likelihood. 2011

338. Wilczek H, Sobra J, Ceska R, Justova V, Juzova Z, Prochazkova R, Kvasilova M. [monitoring plasma levels of vitamin d metabolites in simvastatin (zocor) therapy in patients with familial hypercholesterolemia]. *Casopis lekaru ceskych*. 1994;133:727-729

339. Yavuz B, Ertugrul DT, Cil H, Ata N, Akin KO, Yalcin AA, Kucukazman M, Dal K, Hokkaomeroglu MS, Yavuz BB, Tutal E. Increased levels of 25 hydroxyvitamin d and 1,25-dihydroxyvitamin d after rosuvastatin treatment: A novel pleiotropic effect of statins? *Cardiovascular drugs and therapy / sponsored by the International Society of Cardiovascular Pharmacotherapy*. 2009;23:295-299

340. Perez-Castrillon JL, Abad Manteca L, Vega G, Del Pino Montes J, de Luis D, Duenas Laita A. Vitamin d levels and lipid response to atorvastatin. *International journal of endocrinology*. 2010;2010:320721

341. Freathy RM, Timpson NJ, Lawlor DA, Pouta A, Ben-Shlomo Y, Ruokonen A, Ebrahim S, Shields B, Zeggini E, Weedon MN, Lindgren CM, Lango H, Melzer D, Ferrucci L, Paolisso G, Neville MJ, Karpe F, Palmer CN, Morris AD, Elliott P, Jarvelin MR, Smith GD, McCarthy MI, Hattersley AT, Frayling TM. Common variation in the fto gene alters diabetes-related metabolic traits to the extent expected given its effect on bmi. *Diabetes*. 2008;57:1419-1426

342. Demidenko E. Sample size determination for logistic regression revisited. *Stat Med*. 2007;26:3385-3397

343. Elamin MB, Abu Elnour NO, Elamin KB, Fatourechi MM, Alkatib AA, Almandoz JP, Liu H, Lane MA, Mullan RJ, Hazem A, Erwin PJ, Hensrud DD, Murad MH, Montori VM. Vitamin d and cardiovascular outcomes: A systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2011;96:1931-1942

344. Trivedi DP, Doll R, Khaw KT. Effect of four monthly oral vitamin d3 (cholecalciferol) supplementation on fractures and mortality in men and women living in the community: Randomised double blind controlled trial. *BMJ*. 2003;326:469

345. Bolland MJ, Avenell A, Baron JA, Grey A, MacLennan GS, Gamble GD, Reid IR. Effect of calcium supplements on risk of myocardial infarction and cardiovascular events: Meta-analysis. *BMJ*. 2010;341:c3691

346. Vacek JL, Vanga SR, Good M, Lai SM, Lakkireddy D, Howard PA. Vitamin d deficiency and supplementation and relation to cardiovascular health. *Am J Cardiol*. 2012;109:359-363

347. Muller K, Bendtzen K. 1,25-dihydroxyvitamin d3 as a natural regulator of human immune functions. *J.Investig.Dermatol.Symp.Proc*. 1996;1:68-71

348. Sugden JA, Davies JI, Witham MD, Morris AD, Struthers AD. Vitamin d improves endothelial function in patients with type 2 diabetes mellitus and low vitamin d levels. *Diabet Med*. 2008;25:320-325

349. Watson KE, Abrolat ML, Malone LL, Hoeg JM, Doherty T, Detrano R, Demer LL. Active serum vitamin d levels are inversely correlated with coronary calcification. *Circulation*. 1997;96:1755-1760

350. Ishii M, Tanabe Y, Goto M, Sugita K. [mri as an aid for diagnosis of infantile neuroaxonal dystrophy]. *No to hattatsu. Brain and development*. 1992;24:491-493

351. Heliovaara M, Karvonen MJ, Vilhunen R, Punsar S. Smoking, carbon monoxide, and atherosclerotic diseases. *Br Med J*. 1978;1:268-270

352. Price JF, Mowbray PI, Lee AJ, Rumley A, Lowe GD, Fowkes FG. Relationship between smoking and cardiovascular risk factors in the development of peripheral arterial disease and coronary artery disease: Edinburgh artery study. *Eur Heart J*. 1999;20:344-353

353. Alzamora MT, Fores R, Baena-Diez JM, Pera G, Toran P, Sorribes M, Vicheto M, Reina MD, Sancho A, Albaladejo C, Llussa J. The peripheral arterial disease study (perart/artper): Prevalence and risk factors in the general population. *BMC public health*. 2010;10:38

354. Fowkes FG, Housley E, Riemersma RA, Macintyre CC, Cawood EH, Prescott RJ, Ruckley CV. Smoking, lipids, glucose intolerance, and blood pressure as risk factors for peripheral atherosclerosis compared with ischemic heart disease in the edinburgh artery study. *Am J Epidemiol*. 1992;135:331-340

355. Donnelly LA, Zuydam NV, Zhou K, Tavendale R, Carr F, Zee A-HM-vd, Leusink M, Boer Ad, Klungel OH, Doevendans PA, Morris AD, Pearson ER, Doney AS, Palmer CN. Robust association of the lpa locus with ldlc lowering response to statin treatment in a meta-analysis of 30,467 individuals from both randomised control trials and observational studies and association with coronary artery disease outcome during statin treatment.

356. Manolio TA, Weis BK, Cowie CC, Hoover RN, Hudson K, Kramer BS, Berg C, Collins R, Ewart W, Gaziano JM, Hirschfeld S, Marcus PM, Masys D, McCarty CA, McLaughlin J, Patel AV, Peakman T, Pedersen NL, Schaefer C, Scott JA, Sprosen T, Walport M, Collins FS. New models for large prospective studies: Is there a better way? *Am J Epidemiol*. 2012;175:859-866

357. Schaefer C, Sciortino S, Kvale M, Lapham K, Lin J, Ranatunga D, Rowell S, Sadler M, Miles S, McGuire W, Ludwig D, Walter L, Listerman I, Eeden SVD, Whitmer R, Quesenberry C, Risch N, Blackburn E. The kaiser permanente/ucsf genetic epidemiology research study on adult health and aging: Demographic and behavioral influences on telomeres and relationship with all-cause mortality. *American Society for Human Genetics*. 2012

358. Anderson CD, Nalls MA, Biffi A, Rost NS, Greenberg SM, Singleton AB, Meschia JF, Rosand J. The effect of survival bias on case-control genetic association studies of highly lethal diseases. *Circ Cardiovasc Genet*. 2011;4:188-196

359. Flex A, Gaetani E, Pola R, Santoliquido A, Aloi F, Papaleo P, Dal Lago A, Pola E, Serricchio M, Tondi P, Pola P. The -174 g/c polymorphism of the interleukin-6 gene promoter is associated with peripheral artery occlusive disease. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery*. 2002;24:264-268

360. Wang KS, Liu X, Zhang Q, Zeng M. Anapc1 and slco3a1 are associated with nicotine dependence: Meta-analysis of genome-wide association studies. *Drug and alcohol dependence*. 2012;124:325-332

361. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag snps identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*. 2008;40:616-622

362. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, Fugman D, Goate AM, Hinrichs AL, Konvicka K, Martin NG, Montgomery GW, Saccone NL, Saccone SF, Wang JC, Chase GA, Rice JP, Ballinger DG. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 2007;16:24-35

363. Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau O, Swan GE, Goate AM, Rutter J, Bertelsen S, Fox L, Fugman D, Martin NG, Montgomery GW, Wang JC, Ballinger DG, Rice JP, Bierut LJ. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 snps. *Hum Mol Genet*. 2007;16:36-49

364. Leeper NJ, Raiesdana A, Kojima Y, Kundu RK, Cheng H, Maegdefessel L, Toh R, Ahn GO, Ali ZA, Anderson DR, Miller CL, Roberts SC, Spin JM, de Almeida PE, Wu JC, Xu B, Cheng K, Quertermous M, Kundu S, Kortekaas KE, Berzin E, Downing KP, Dalman RL, Tsao PS, Schadt EE, Owens GK, Quertermous T. Loss of cdkn2b promotes p53-dependent smooth muscle cell apoptosis and aneurysm formation. *Arterioscler Thromb Vasc Biol*. 2013;33:e1-e10

365. Li YY. Lack of association of a-6g polymorphism of agt gene with essential hypertension in the chinese population. *J Cardiovasc Med (Hagerstown)*. 2012;13:505-510

366. Chanda K, Chou CT, Lai JJ, Lin SF, Yellol GS, Sun CM. Traceless synthesis of diketopiperazine fused tetrahydro-beta-carbolines on soluble polymer support. *Molecular diversity*. 2011;15:569-581

67. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide

association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*. 2008;40:638-645

368. Barber MJ, Mangravite LM, Hyde CL, Chasman DI, Smith JD, McCarty CA, Li X, Wilke RA, Rieder MJ, Williams PT, Ridker PM, Chatterjee A, Rotter JI, Nickerson DA, Stephens M, Krauss RM. Genome-wide association of lipid-lowering response to statins in combined study populations. *PLoS One*. 2010;5:e9763

369. Hung FC, Chang Y, Lin-Chao S, Chao CC. Gas7 mediates the differentiation of human bone marrow-derived mesenchymal stem cells into functional osteoblasts by enhancing runx2-dependent gene expression. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*. 2011;29:1528-1535

370. Baud'huin M, Solban N, Cornwall-Brady M, Sako D, Kawamoto Y, Liharska K, Lath D, Bouxsein ML, Underwood KW, Ucran J, Kumar R, Pobre E, Grinberg A, Seehra J, Canalis E, Pearsall RS, Croucher PI. A soluble bone morphogenetic protein type ia receptor increases bone mass and bone strength. *Proc Natl Acad Sci U S A*. 2012;109:12207-12212

371. Morales J, Al-Sharif L, Khalil DS, Shinwari JM, Bavi P, Al-Mahrouqi RA, Al-Rajhi A, Alkuraya FS, Meyer BF, Al Tassan N. Homozygous mutations in adamts10 and adamts17 cause lenticular myopia, ectopia lentis, glaucoma, spherophakia, and short stature. *Am J Hum Genet*. 2009;85:558-568

372. Talmud PJ, Hawe E, Miller GJ. Analysis of gene-environment interaction in coronary artery disease: Lipoprotein lipase and smoking as examples. *Italian heart journal : official journal of the Italian Federation of Cardiology*. 2002;3:6-9

373. Montasser ME, Shimmin LC, Hanis CL, Boerwinkle E, Hixson JE. Gene by smoking interaction in hypertension: Identification of a major quantitative trait locus on chromosome 15q for systolic blood pressure in mexican-americans. *J Hypertens*. 2009;27:491-501

374. Luke MM, Berger K, Rowland CM, Catanese JJ, Tong CH, Ross DA, Garcia V, Kuhlenbaeumer G, Ringelstein EB, Pullinger CR, Malloy MJ, Deedwania P, Ellis SG, Kane JP, Devlin JJ, Lalouschek W, Mannhalter C. Polymorphisms and noncardioembolic stroke in three case-control studies. *Cerebrovasc Dis*. 2012;33:80-85

375. Nishizaki R, Ota M, Inoko H, Meguro A, Shiota T, Okada E, Mok J, Oka A, Ohno S, Mizuki N. New susceptibility locus for high myopia is linked to the uromodulin-like 1 (umodl1) gene region on chromosome 21q22.3. *Eye (Lond)*. 2009;23:222-229

376. Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, McEvoy BP, Schrage AJ, Grant JD, Chou YL, Zhu R, Henders AK, Medland SE, Gordon SD, Nelson EC, Agrawal A, Nyholt DR, Bucholz KK, Madden PA, Montgomery GW. A quantitative-trait genome-wide association study of alcoholism risk in the community: Findings and implications. *Biological psychiatry*. 2011;70:513-518

377. Pare G, Chasman DI, Parker AN, Nathan DM, Miletich JP, Zee RY, Ridker PM. Novel association of hk1 with glycated hemoglobin in a non-diabetic population: A genome-wide

evaluation of 14,618 participants in the women's genome health study. *PLoS Genet*. 2008;4:e1000312

378. Flammer AJ, Gossl M, Li J, Matsuo Y, Reriani M, Loeffler D, Simari RD, Lerman LO, Khosla S, Lerman A. Patients with an hba1c in the prediabetic and diabetic range have higher numbers of circulating cells with osteogenic and endothelial progenitor cell markers. *J Clin Endocrinol Metab*. 2012;97:4761-4768

379. Yasuno K, Bilguvar K, Bijlenga P, Low SK, Krischek B, Auburger G, Simon M, Krex D, Arlier Z, Nayak N, Ruigrok YM, Niemela M, Tajima A, von und zu Fraunberg M, Doczi T, Wirjatijasa F, Hata A, Blasco J, Oszvald A, Kasuya H, Zilani G, Schoch B, Singh P, Stuer C, Risselada R, Beck J, Sola T, Ricciardi F, Aromaa A, Illig T, Schreiber S, van Duijn CM, van den Berg LH, Perret C, Proust C, Roder C, Ozturk AK, Gaal E, Berg D, Geisen C, Friedrich CM, Summers P, Frangi AF, State MW, Wichmann HE, Breteler MM, Wijmenga C, Mane S, Peltonen L, Elio V, Sturkenboom MC, Lawford P, Byrne J, Macho J, Sandalcioglu EI, Meyer B, Raabe A, Steinmetz H, Rufenacht D, Jaaskelainen JE, Hernesniemi J, Rinkel GJ, Zembutsu H, Inoue I, Palotie A, Cambien F, Nakamura Y, Lifton RP, Gunel M. Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat Genet*. 2010;42:420-425

380. Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stancakova A, Stringham HM, Sim X, Yang L, Fuchsberger C, Cederberg H, Chines PS, Teslovich TM, Romm JM, Ling H, McMullen I, Ingersoll R, Pugh EW, Doheny KF, Neale BM, Daly MJ, Kuusisto J, Scott LJ, Kang HM, Collins FS, Abecasis GR, Watanabe RM, Boehnke M, Laakso M, Mohlke KL. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet*. 2012

381. Ordovas JM, Smith CE. Epigenetics and cardiovascular disease. *Nature reviews. Cardiology*. 2010;7:510-519

382. Tsai CT, Hwang JJ, Ritchie MD, Moore JH, Chiang FT, Lai LP, Hsu KL, Tseng CD, Lin JL, Tseng YZ. Renin-angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort: Detection of high order gene-gene interaction. *Atherosclerosis*. 2007;195:172-180

383. Bi M, Kao WH, Boerwinkle E, Hoogeveen RC, Rasmussen-Torvik LJ, Astor BC, North KE, Coresh J, Kottgen A. Association of rs780094 in gckr with metabolic traits and incident diabetes and cardiovascular disease: The aric study. *PLoS One*. 2010;5:e11690

384. Davies RW, Wells GA, Stewart AF, Erdmann J, Shah SH, Ferguson JF, Hall AS, Anand SS, Burnett MS, Epstein SE, Dandona S, Chen L, Nahrstaedt J, Loley C, Konig IR, Kraus WE, Granger CB, Engert JC, Hengstenberg C, Wichmann HE, Schreiber S, Tang WH, Ellis SG, Rader DJ, Hazen SL, Reilly MP, Samani NJ, Schunkert H, Roberts R, McPherson R. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet*. 2012;5:217-225

385. Wild PS, Zeller T, Schillert A, Szymczak S, Sinning CR, Deiseroth A, Schnabel RB, Lubos E, Keller T, Eleftheriadis MS, Bickel C, Rupprecht HJ, Wilde S, Rossmann H, Diemert P, Cupples LA, Perret C, Erdmann J, Stark K, Kleber ME, Epstein SE, Voight BF, Kuulasmaa K, Li M, Schafer AS, Klopp N, Braund PS, Sager HB, Demissie S, Proust C, Konig IR, Wichmann HE, Reinhard W, Hoffmann MM, Virtamo J, Burnett MS, Siscovick D, Wiklund PG, Qu L, El

Mokthari NE, Thompson JR, Peters A, Smith AV, Yon E, Baumert J, Hengstenberg C, Marz W, Amouyel P, Devaney J, Schwartz SM, Saarela O, Mehta NN, Rubin D, Silander K, Hall AS, Ferrieres J, Harris TB, Melander O, Kee F, Hakonarson H, Schrezenmeir J, Gudnason V, Elosua R, Arveiler D, Evans A, Rader DJ, Illig T, Schreiber S, Bis JC, Altshuler D, Kavousi M, Witteman JC, Uitterlinden AG, Hofman A, Folsom AR, Barbalic M, Boerwinkle E, Kathiresan S, Reilly MP, O'Donnell CJ, Samani NJ, Schunkert H, Cambien F, Lackner KJ, Tiret L, Salomaa V, Munzel T, Ziegler A, Blankenberg S. A genome-wide association study identifies lipa as a susceptibility gene for coronary artery disease. *Circ Cardiovasc Genet*. 2011;4:403-412

386. Slavin TP, Feng T, Schnell A, Zhu X, Elston RC. Two-marker association tests yield new disease associations for coronary artery disease and hypertension. *Human genetics*. 2011;130:725-733

387. Erdmann J, Willenborg C, Nahrstaedt J, Preuss M, Konig IR, Baumert J, Linsel-Nitschke P, Gieger C, Tennstedt S, Belcredi P, Aherrahrou Z, Klopp N, Loley C, Stark K, Hengstenberg C, Bruse P, Freyer J, Wagner AK, Medack A, Lieb W, Grosshennig A, Sager HB, Reinhardt A, Schafer A, Schreiber S, El Mokhtari NE, Raaz-Schrauder D, Illig T, Garlichs CD, Ekici AB, Reis A, Schrezenmeir J, Rubin D, Ziegler A, Wichmann HE, Doering A, Meisinger C, Meitinger T, Peters A, Schunkert H. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur Heart J*. 2011;32:158-168

Appendices

Appendix 1: SNPs and platforms used to calculate the genotype risk scores for type 1 diabetes in the Go-DARTS study

CHR	BP	Gene	SNP	Immuno Chip SNP	Affymetrix 6.0 Illumina Dual SNP
1	64108771	PGM1	rs2269241	rs2269241	rs2269241
1	1.14E+08	PTPN22	rs2476601	rs2476601	rs1230666
1	1.93E+08	RGS21	rs2816316	rs1323296	rs1323296
1	2.07E+08	MAPKAPK2	rs3024505	rs3024505	rs3024505
2	1.03E+08	IL18RAP	rs917997	rs7559479	rs7559479
2	1.63E+08	IFIH1	rs1990760	rs7608315	rs7608315
2	2.05E+08	CTLA4	rs3087243	rs3087243	rs3087243
3	46345611	CCR3	rs11711054	rs6441961	rs6441961
4	26085511	C4orf52	rs10517086	rs932036	rs932036
4	1.23E+08	KIAA1109	rs4505848	rs7671357	rs7671357
5	35874575	IL7R	rs6897932	rs6897932	rs6897932
6	32408527	HLA-DRA	rs9268645	rs9268645	rs9268645
6	90958231	BACH2	rs11755527	rs4707605	rs4707605
6	1.38E+08	BTF3L4P3	rs2327832	rs2327832	rs2327832
7	26891665	SKAP2	rs7804356	rs10486483	rs10486483
7	51027194	GRB10	rs4948088	rs4948088	rs4948088
9	4291747	GLIS3	rs7020673	rs7020673	rs7020673
10	6123495	IL2RA	rs12251307	rs12251307	rs12251307
10	6472891	PRKCQ	rs11258747	rs11258747	rs11258747
10	90023033	PTEN	rs10509540	rs7068821	rs7068821
11	2213166	MIR4686	rs7111341	rs11564708	rs11564708
12	9910164	CD69	rs4763879	rs4763879	rs4763879
12	56482180	ERBB3	rs2292239	rs2292239	rs2292239
12	1.12E+08	SH2B3	rs3184504	rs653178	rs653178
14	69263599	ZFP36L1	rs1465788	rs1465788	rs1465788
14	98498951	C14orf64	rs4900384	rs4900384	rs4900384
15	79235446	CTSH	rs3825932	rs3825932	rs3825932
16	11179873	CLEC16A	rs12708716	rs12708716	rs12708716
16	28539848	IL27	rs4788084	rs1074631	rs1074631
16	75247245	CTRB2	rs7202877	rs4993971	rs4993971
17	7633692	DNAH2	rs16956936	rs16956936	rs16956936
17	38066240	GSDMB	rs2290400	rs1008723	rs1008723
17	38770286	CCR7	rs7221109	rs7221109	rs7221109
18	12809340	PTPN2	rs1893217	rs2542151	rs2542151
20	1610551	SIRPG	rs2281808	rs2281808	rs2281808
21	43836186	UBASH3A	rs11203203	rs11203203	rs11203203
22	30581722	HORMAD2	rs5753037	rs4820830	rs4820830
22	37591318	C1QTNF6	rs229541	rs229541	rs229541

Appendix 2: SNPs and platforms used to calculate the genotype risk score for type 2 diabetes in the Go-DARTS study

CHR	BP	Gene	SNP	CardioMetabochip Chip SNP	Affymetrix 6.0 Illumina Dual SNP
1	120517959	NOTCH2	rs10923931	rs10923931	rs10923931
1	214154719	PROX1-AS1	rs2075423	rs2075423	rs2075423
2	27741237	GCKR	rs780094	rs780094	rs780094
2	43690030	THADA	rs10203174	rs10203174	rs10203174
2	60568745	RNA5SP94	rs243088	rs243088	rs243090
2	161346447	RBMS1	rs7569522	rs7569522	rs7569522
2	165528876	COBLL1	rs13389219	rs6717858	rs6717858
2	227093585	NYAP2	rs2943640	rs2943640	rs2943640
3	12393125	PPARG	rs1801282	rs1801282	rs1801282
3	23454790	UBE2E2	rs1496653	rs1496653	rs1496653
3	64090363	PRICKLE2-AS2	rs12497268	rs12497268	rs12497268
3	64705365	ADAMTS9-AS2	rs6795735	rs6795735	rs6795735
3	123082398	ADCY5	rs11717195	rs11717195	rs11717195
3	185511687	IGF2BP2	rs4402960	rs4402960	rs4402960
3	186613409	ADIPOQ	rs17301514	rs17301514	rs7648806
4	6289986	WFS1	rs4458523	rs4458523	rs4458523
5	55806751	HMGN1P17	rs459193	rs40271	rs40271
6	20679709	CDKAL1	rs7756992	rs7756992	rs7756992
6	38177667	BTBD9	rs4299828	rs4299828	rs4299828
6	39304211	KIF6	rs3734621	rs3734621	rs3734621
7	14898282	DGKB	rs17168486	rs17168486	rs17168486
7	28196413	JAZF1	rs849135	rs864745	rs864745
7	130437689	KLF14	rs13233731	rs13234407	rs13233731
8	41519248	ANK1	rs516946	rs516946	rs516946
8	95937502	NDUFAF6	rs7845219	rs7845219	rs7845219
8	118185025	SLC30A8	rs3802177	rs3802177	rs3802177
9	4292083	GLIS3	rs10758593	rs4237150	rs4237150
9	8369533	PTPRD	rs16927668	rs16927668	rs16927668
9	22051670	CDKN2B-AS1	rs944801	rs7030641	rs944801
9	22134094	CDKN2B-AS1	rs10811661	rs10811661	rs10811661
9	81905590	MTND2P8	rs17791513	rs17791513	rs13292347
9	84308948	TLE1	rs2796441	rs2796441	rs2796441
10	12307894	CDC123	rs11257655	rs4747969	rs4747969
10	70865342	SRGN	rs12242953	rs12242953	rs12242953
10	80942631	ZMIZ1	rs12571751	rs12571751	rs12571751
10	94462882	HHEX	rs1111875	rs1111875	rs1111875
10	114758349	TCF7L2	rs7903146	rs4506565	rs4506565
11	1696849	FAM99A	rs2334499	rs2334499	rs4752781

CHR	BP	Gene	SNP	CardioMetabochip Chip SNP	Affymetrix 6.0 Illumina Dual SNP
11	2691500	KCNQ1	rs231361	rs463924	rs463924
11	2847069	KCNQ1	rs163184	rs163177	rs163177
11	17408630	KCNJ11	rs5215	rs5215	rs5215
11	72433098	ARAP1	rs1552224	rs17244499	rs17244499
11	92708710	MTNR1B	rs10830963	rs10830963	rs10830963
12	4374373	PARP11	rs11063069	rs11063069	rs11063069
12	27965150	KLHDC5	rs10842994	rs10842994	rs10842994
12	66212318	RPSAP52	rs2261181	rs2261181	rs2261181
12	71433293	PTPRR	rs7955901	rs7138300	rs7138300
12	121426901	HNF1A	rs12427353	rs7965349	rs7965349
13	80717156	LINC00329	rs1359790	rs1359790	rs1359790
15	62383155	C2CD4A	rs4502156	rs4502156	rs6494307
15	77832762	HMG20A	rs7177055	rs7177055	rs7177055
15	80432222	ZFAND6	rs11634397	rs11634397	rs11634397
15	90345335	ANPEP	rs2007084	rs2007084	rs17240268
15	91544076	VPS33B	rs12899811	rs11073964	rs11073964
16	53819169	FTO	rs9936385	rs8050136	rs8050136
16	75247245	CTRB2	rs7202877	rs3743614	rs3743614
17	36102381	HNF1B	rs11651052	rs11651755	rs11651755
18	57884750	RPS3AP49	rs12970134	rs12970134	rs12970134
19	19407718	SUGP1	rs10401969	rs10401969	rs10401969
19	46158513	EML2	rs8108269	rs8108269	rs8108269
20	42989267	HNF4A	rs4812829	rs4812829	rs16988991

Appendix 3: SNPs and platforms used to calculate the genotype risk score for fasting glucose in the Go-DARTS study

CHR	BP	Gene	SNP	CardioMetabo Chip SNP	Affymetrix.6.0 Illumina.Dual.SNP
1	214159256	PROX1-AS1	rs340874	rs340874	rs340874
2	27741237	GCKR	rs780094	rs780094	rs780094
2	169763148	SPC25	rs560887	rs560887	rs560887
3	123065778	ADCY5	rs11708067	rs11708067	rs11708067
3	170717521	SLC2A2	rs11920090	rs11924648	rs11924648
7	15064309	DGKB	rs2191349	rs4719433	rs4719433
7	44235668	GCK	rs4607517	rs2908289	rs2908289
8	118184783	SLC30A8	rs13266634	rs13266634	rs13266634
9	4289050	GLIS3	rs7034200	rs7024686	rs7024686
10	113042093	ADRA2A	rs10885122	rs4918635	rs4918635
10	114758349	TCF7L2	rs7903146	rs7903146	rs7903146
11	45873091	CRY2	rs11605924	rs6485644	rs11605924
11	47336320	MADD	rs7944584	rs7944584	rs7944584
11	61571478	FADS1	rs174550	rs174577	rs174577
11	92708710	MTNR1B	rs10830963	rs10830963	rs10830963
12	102875569	IGF1	rs35767	rs35767	rs35767
15	62433962	C2CD4A	rs11071657	rs12440695	rs12440695

Appendix 4: SNPs and platforms used to calculate the genetic risk scores for blood pressure in the Go-DARTS study

CHR	BP	Gene	SNP	CardioMetabo Chip SNP	Affymetrix.6.0 Illumina.Dual.SNP
1	11862778	MTHFR	rs17367504	rs17367504	rs17367504
1	113216543	MOV10	rs2932538	rs2932538	rs4839257
3	27537909	SLC4A7	rs13082711	rs13082711	rs17682751
3	41877414	ULK4	rs3774372	rs6599176	rs6599176
3	169100886	MECOM	rs419076	rs223102	rs448378
4	81164723	PRDM8	rs1458038	rs1458038	rs1458038
4	103188709	SLC39A8	rs13107325	rs13107325	rs13107325
4	156645513	GUCY1A3	rs13139571	rs7698460	rs7698460
5	32815028	NPR3	rs1173771	rs1173771	rs1173771
5	157845402	MARK2P11	rs11953630	rs12187017	rs12187017
6	26091179	HFE	rs1799945	rs198846	rs198846
6	31616366	BAG6	rs805303	rs805303	rs805303
10	18419972	SLC39A12	rs4373814	rs12570727	rs12570727
10	18707448	CACNB2	rs1813353	rs11014171	rs11014171
10	63467553	C10orf107	rs4590817	rs2393833	rs2393833
10	95895940	PLCE1	rs932764	rs10786152	rs10786152
10	104846178	NT5C2	rs11191548	rs11191560	rs11191560
11	10350538	AMPD3	rs7129220	rs7129220	rs7129220
11	16902268	PLEKHA7	rs381815	rs11024074	rs11024074
11	100593538	ARHGAP42	rs633185	rs604723	rs604723
12	90060586	ATP2B1	rs17249754	rs2681472	rs2681472
12	111884608	SH2B3	rs3184504	rs4766578	rs3184504
12	113872179	TBX3	rs10850411s	rs10850411	rs10850411
15	75077367	CSK	rs1378942	rs1378942	rs1378942
15	91437388	FES	rs2521501	rs4932370	rs4932370
17	45013271	GOSR2	rs17608766	rs17608766	rs17608766
17	47402807	ZNF652	rs12940887	rs12940887	rs12940887
20	10969030	FAT1P1	rs1327235	rs1327235	rs1327235
20	57751117	MRPS16P	rs6015450	rs16982520	rs16982520

Appendix 5: SNPs and platforms used to calculate genotype risk scores for cholesterol, low density lipoprotein, high density lipoprotein and triglycerides in the Go-DARTS study.

Gene	CHR	BP	Lead.SNP	Trait	Affymetrix .Illumina.SNP	CardioMetabo .Chip
TMEM57	1	25775733	rs12027135	LDL	rs873308	rs873308
TMEM57	1	25775733	rs12027135	TC	rs873308	rs873308
PPIEL	1	40018509	rs4660808	TG	rs17264866	rs17513135
PABPC4	1	40028180	rs4660293	HDL	rs4660293	rs4660293
BSND	1	55504650	rs2479409	LDL	rs2479409	rs2479409
BSND	1	55504650	rs2479409	TC	rs2479409	rs2479409
DOCK7	1	63025942	rs2131925	TC	rs1748195	rs1748195
DOCK7	1	63025942	rs2131925	TG	rs1748195	rs1748195
DOCK7	1	63025942	rs2131925	LDL	rs1748195	rs1748195
EVI5	1	93009438	rs7515577	TC	rs4970712	rs4970712
CCDC18	1	93700212	rs531514	HDL	rs12118262	rs12118262
CELSR2	1	109818306	rs629301	LDL	rs629301	rs629301
CELSR2	1	109818306	rs629301	TC	rs629301	rs629301
ZNF648	1	182168885	rs1689800	HDL	rs2243976	rs2243976
	1	220973563	rs2642442	TC	rs2807834	rs2807834
	1	220973563	rs2642442	LDL	rs2807834	rs2807834
GALNT2	1	230295691	rs4846914	HDL	rs4846914	rs4846914
GALNT2	1	230295691	rs4846914	TG	rs4846914	rs4846914
LINC00184	1	234858597	rs514230	LDL	rs553427	rs558971
LINC00184	1	234858597	rs514230	TC	rs684818	rs558971
APOB	2	21225281	rs1042034	TG	rs673548	rs673548
APOB	2	21225281	rs1042034	HDL	rs673548	rs673548
APOB	2	21263900	rs1367117	LDL	rs1367117	rs1367117
APOB	2	21263900	rs1367117	TC	rs1367117	rs1367117
APOB	2	21286057	rs515135	LDL	rs515135	rs515135
APOB	2	21286057	rs515135	TC	rs515135	rs515135
APOB	2	21291529	rs668948	TG	rs488507	rs488507
GCKR	2	27730940	rs1260326	TG	rs1260326	rs1260326
GCKR	2	27730940	rs1260326	TC	rs1260326	rs1260326
ABCG8	2	44072576	rs4299376	LDL	rs6544713	rs6544713
ABCG8	2	44072576	rs4299376	TC	rs6544713	rs6544713
ABCG8	2	44074000	rs4953023	TC	rs4953023	rs4953023
ABCG8	2	44074000	rs4953023	LDL	rs4953023	rs4953023
RAB3GAP1	2	135837906	rs7570971	TC	rs6730157	rs6730157
RAB3GAP1	2	135893372	rs10445686	LDL	rs16831243	rs16831243
COBLL1	2	165513091	rs10195252	TG	rs10195252	rs10195252
COBLL1	2	165540800	rs12328675	HDL	rs7607980	rs7607980
NYAP2	2	227099180	rs2943645	TG	rs2972147	rs2972147

Gene	CHR	BP	Lead.SNP	Trait	Affymetrix .Illumina.SNP	CardioMetabo .Chip
NYAP2	2	227100698	rs2972146	HDL	rs2972147	rs2972147
RAF1	3	12628920	rs2290159	TC	rs7956	rs11709504
MSL2	3	135926622	rs645040	TG	rs645040	rs645040
AFF1	4	88030261	rs442177	TG	rs442177	rs442177
AFF1	4	88030261	rs442177	HDL	rs442177	rs442177
SLC39A8	4	103188709	rs13107325	HDL	rs13107325	rs13107325
ARL15	5	53298025	rs6450176	HDL	rs6876198	rs4311394
HMGN1P17	5	55861786	rs9686661	TG	rs9686661	rs9686661
HMGCR	5	74656539	rs12916	LDL	rs12916	rs12916
HMGCR	5	74656539	rs12916	TC	rs12916	rs12916
TIMD4	5	156390297	rs6882076	TC	rs1501908	rs1501908
TIMD4	5	156390297	rs6882076	LDL	rs1501908	rs1501908
HAVCR1	5	156479323	rs1553318	TG	rs7724832	rs7724832
ARPC3P5	6	16127407	rs3757354	LDL	rs3757354	rs3757354
ARPC3P5	6	16127407	rs3757354	TC	rs3757354	rs3757354
HFE	6	26093141	rs1800562	LDL	rs1800562	rs1800562
HFE	6	26093141	rs1800562	TC	rs1800562	rs1800562
WASF5P	6	31265490	rs2247056	TG	rs9264602	rs2247056
HLA-DRA	6	32412435	rs3177928	TC	rs3177928	rs3177928
HLA-DRA	6	32412435	rs3177928	LDL	rs3177928	rs3177928
HLA-DRA	6	32683961	rs12660719	TG	rs12660719	rs12660719
SPDEF	6	34546560	rs2814982	TC	rs2814982	rs2814982
SPDEF	6	34552797	rs2814944	HDL	rs2814944	rs2814944
TCP11	6	35133074	rs3800406	LDL	rs3800406	rs3800406
FRK	6	116312893	rs9488822	TC	rs3798236	rs3798236
FRK	6	116354591	rs11153594	LDL	rs11153594	rs11153594
CITED2	6	139829666	rs605066	HDL	rs634869	rs634869
CITED2	6	139843583	rs636202	TG	rs628751	rs628751
SLC22A1	6	160578860	rs1564348	LDL	rs1564348	rs1564348
SLC22A1	6	160578860	rs1564348	TC	rs1564348	rs1564348
SLC22A3	6	160774441	rs486359	TG	rs487060	rs501470
LPA	6	161010118	rs10455872	LDL	rs10455872	rs10455872
LPA	6	161010118	rs10455872	TC	rs10455872	rs10455872
LPA	6	161089817	rs1084651	HDL	rs1084651	rs783149
DNAH11	7	21607352	rs12670798	LDL	rs12670798	rs12670798
DNAH11	7	21607352	rs12670798	TC	rs5008148	rs5008148
NPC1L1	7	44579180	rs2072183	TC	rs17725246	rs17725246
NPC1L1	7	44600695	rs217386	LDL	rs217381	rs217373
BAZ1B	7	72934510	rs7811265	TG	rs1178977	rs11974409
BCL7B	7	72982874	rs17145738	HDL	rs17145738	rs17145738
KLF14	7	130433384	rs4731702	HDL	rs11979110	rs11979110
KLF14	7	130457931	rs1562398	TG	rs1596972	rs1596972
PPP1R3B	8	9183358	rs9987289	HDL	rs9987289	rs2126259
PPP1R3B	8	9185146	rs2126259	LDL	rs1461729	rs2126259

Gene	CHR	BP	Lead.SNP	Trait	Affymetrix .Illumina.SNP	CardioMetabo .Chip
SOX7	8	10683929	rs11776767	TG	rs7000939	rs7000939
NAT2	8	18272881	rs1495741	TC	rs1495743	rs4921914
NAT2	8	18272881	rs1495741	TG	rs1495743	rs4921914
LPL	8	19806631	rs7016529	HDL	rs6586879	rs6586879
LPL	8	19806631	rs7016529	TG	rs1031045	rs7016529
LPL	8	19844222	rs12678919	TG	rs12682115	rs12678919
LPL	8	19844222	rs12678919	HDL	rs12682115	rs12678919
FAM110B	8	59311697	rs1030431	LDL	rs13277801	rs13277801
UBXN2B	8	59388565	rs2081687	TC	rs13277801	rs13277801
TRPS1	8	116599199	rs2293889	HDL	rs4599845	rs4599845
TRPS1	8	116648565	rs2737229	TC	rs2737229	rs2737252
TRIB1	8	126490972	rs2954029	TG	rs2954029	rs2954029
TRIB1	8	126490972	rs2954029	LDL	rs2954029	rs2954029
TRIB1	8	126490972	rs2954029	TC	rs2954029	rs2954029
TRIB1	8	126495818	rs10808546	HDL	rs2954029	rs2954029
TRIB1	8	126502526	rs12677676	TC	rs8180991	rs8180991
PLEC	8	145043543	rs11136341	LDL	rs7832643	rs7832643
PLEC	8	145043543	rs11136341	TC	rs7832643	rs7832643
TTC39B	9	15296034	rs643531	HDL	rs643531	rs643531
TTC39B	9	15305378	rs581080	TC	rs581080	rs585002
ABCA1	9	107588777	rs4149311	TC	rs4149311	rs4149311
ABCA1	9	107664301	rs1883025	HDL	rs1883025	rs1883025
ABCA1	9	107664301	rs1883025	TC	rs1883025	rs1883025
ABCA1	9	107665978	rs1800978	LDL	rs1800978	rs1800978
LCN1P1	9	135144821	rs9411489	TC	rs651007	rs579459
ABO	9	136154304	rs649129	LDL	rs651007	rs579459
JMJD1C	10	65027610	rs10761731	TG	rs10761741	rs10761741
CYP26A1	10	94839642	rs2068888	TG	rs2068888	rs2068888
GPAM	10	113910721	rs1129555	LDL	rs1129555	rs1129555
GPAM	10	113933886	rs2255141	TC	rs1129555	rs2255141
AMPD3	11	10388782	rs2923084	HDL	rs2923084	rs2923084
SPTY2D1	11	18632984	rs10128711	LDL	rs10128711	rs10128711
SPTY2D1	11	18632984	rs10128711	TC	rs10128711	rs10128711
F2	11	46743247	rs3136441	HDL	rs2290883	rs2290883
FADS1	11	61569830	rs174546	LDL	rs174577	rs174577
FADS1	11	61569830	rs174546	TC	rs174577	rs174577
FADS1	11	61569830	rs174546	TG	rs174577	rs174577
FADS2	11	61623140	rs174601	HDL	rs174577	rs174577
ZNF259	11	116648917	rs964184	HDL	rs964184	rs6589564
ZNF259	11	116648917	rs964184	TC	rs964184	rs6589564
ZNF259	11	116648917	rs964184	TG	rs964184	rs6589564
ZNF259	11	116648917	rs964184	LDL	rs964184	rs6589564
APOA5	11	116665079	rs9804646	TC	rs9804646	rs9804646
SIK3	11	116728630	rs12225230	HDL	rs11216162	rs11216162

Gene	CHR	BP	Lead.SNP	Trait	Affymetrix .Illumina.SNP	CardioMetabo .Chip
GLULP3	11	122522375	rs7941030	TC	rs7941030	rs7941030
UBASH3B	11	122530591	rs7115089	HDL	rs7941030	rs7941030
ST3GAL4	11	126243952	rs11220462	LDL	rs10893498	rs10893498
ST3GAL4	11	126243952	rs11220462	TC	rs10893498	rs10893498
AEBP2	12	20473758	rs7134375	HDL	rs7134375	rs7134375
R3HDM2	12	57792580	rs11613352	HDL	rs11613352	rs11613352
R3HDM2	12	57792580	rs11613352	TG	rs11613352	rs11613352
MMAB	12	110000193	rs7134594	HDL	rs888192	rs888192
ATXN2	12	112072424	rs11065987	LDL	rs11065987	rs11065987
ATXN2	12	112072424	rs11065987	TC	rs11065987	rs11065987
HNF1A-AS1	12	121416650	rs1169288	TC	rs2650000	rs2650000
HNF1A-AS1	12	121416650	rs1169288	LDL	rs2650000	rs2650000
RNA5SP375	12	123773263	rs11057244	HDL	rs10846513	rs10846513
SBNO1	12	123796238	rs4759375	HDL	rs10773003	rs10846509
ZNF664	12	124460167	rs4765127	HDL	rs11057408	rs11057408
ZNF664	12	124486678	rs12310367	TG	rs7975482	rs7975482
NCOR2	12	125261593	rs838880	HDL	rs838880	rs838880
NYNRIN	14	24872209	rs6573778	TC	rs6573778	rs6573778
NYNRIN	14	24883887	rs8017377	LDL	rs8017377	rs8017377
CAPN3	15	42683787	rs2412710	TG	rs2412710	rs2412710
FRMD5	15	44245931	rs2929282	TG	rs492571	rs492571
ALDH1A2	15	58683366	rs1532085	HDL	rs1532085	rs1532085
ALDH1A2	15	58683366	rs1532085	TC	rs1532085	rs1532085
ALDH1A2	15	58723939	rs2070895	HDL	rs2070895	rs1077834
ALDH1A2	15	58726744	rs261334	TG	rs261332	rs588136
ALDH1A2	15	58731153	rs261342	TC	rs261342	rs588136
ALDH1A2	15	58731153	rs261342	TG	rs261342	rs588136
TPM1	15	63396867	rs2652834	HDL	rs2729783	rs2652834
CTF2P	16	30918487	rs11649653	TG	rs11649653	rs4889606
HERPUD1	16	56989590	rs247616	LDL	rs247617	rs247617
HERPUD1	16	56993324	rs3764261	TC	rs247617	rs247617
HERPUD1	16	56993324	rs3764261	HDL	rs247617	rs247617
CETP	16	57002732	rs9939224	HDL	rs9939224	rs7499892
CETP	16	57004889	rs7205804	TG	rs1532625	rs1532625
PSKH1	16	67928042	rs16942887	HDL	rs1124324	rs1124324
HPR	16	72108093	rs2000999	TC	rs2000999	rs2000999
HPR	16	72108093	rs2000999	LDL	rs2000999	rs2000999
CMIP	16	81534790	rs2925979	HDL	rs2925979	rs2925979
STARD3	17	37813856	rs11869286	HDL	rs931992	rs931992
ITGB3	17	45391804	rs7225700	LDL	rs7225700	rs7225700
EFCAB13	17	45425115	rs7206971	TC	rs12452315	rs12452315
ABCA8	17	66875294	rs4148008	HDL	rs4148005	rs4148005
PGS1	17	76403984	rs4129767	HDL	rs2376583	rs4366775
LIPG	18	47160953	rs7241918	HDL	rs7240405	rs7240405

Gene	CHR	BP	Lead.SNP	Trait	Affymetrix .Illumina.SNP	CardioMetabo .Chip
LIPG	18	47160953	rs7241918	TC	rs7240405	rs7240405
LIPG	18	47278345	rs2040293	HDL	rs7241641	rs7241641
RPS3AP49	18	57849023	rs12967135	HDL	rs476828	rs476828
ANGPTL4	19	8433196	rs7255436	HDL	rs2278236	rs2278236
LDLR	19	11202306	rs6511720	LDL	rs6511720	rs6511720
LDLR	19	11202306	rs6511720	TC	rs6511720	rs6511720
LDLR	19	11224265	rs5930	LDL	rs5930	rs5930
LDLR	19	11227602	rs688	TC	rs688	rs688
DOCK6	19	11347493	rs737337	HDL	rs737337	rs737337
SUGP1	19	19407718	rs10401969	LDL	rs10401969	rs10401969
SUGP1	19	19407718	rs10401969	TG	rs10401969	rs10401969
SUGP1	19	19407718	rs10401969	TC	rs10401969	rs10401969
APOE	19	45414451	rs439401	TG	rs439401	rs439401
APOE	19	45415640	rs445925	LDL	rs7412	rs7412
APOC1	19	45422946	rs4420638	HDL	rs4420638	rs4420638
APOC1	19	45422946	rs4420638	LDL	rs4420638	rs4420638
APOC1	19	45422946	rs4420638	TC	rs4420638	rs4420638
APOC4	19	45448465	rs5167	HDL	rs5167	rs5167
FUT2	19	49206417	rs492602	LDL	rs492602	rs492602
FUT2	19	49206417	rs492602	TC	rs516316	rs516246
MIR4752	19	54792761	rs386000	HDL	rs103294	rs103294
FER1L4	20	34152782	rs2277862	TC	rs2277862	rs2277862
ATG3P1	20	39091487	rs2902940	TC	rs2902940	rs2902940
ATG3P1	20	39091514	rs2902941	LDL	rs2902940	rs2902941
ATG3P1	20	39181246	rs6016382	LDL	rs6016382	rs6016381
ATG3P1	20	39181246	rs6016382	TC	rs6016382	rs6016381
TOP1	20	39672618	rs6029526	TC	rs760762	rs6065311
ZHX3	20	39936815	rs909802	LDL	rs6029609	rs6029609
HNF4A	20	43042364	rs1800961	HDL	rs1800961	rs1800961
HNF4A	20	43042364	rs1800961	TC	rs1800961	rs1800961
PLTP	20	44554015	rs6065906	HDL	rs7679	rs7679

Appendix 6: Table of SNPs that have been previously associated with coronary artery disease that were used in the genotypic risk score calculation

Gene	CHR	BP	SNP	Affymetrix.Illumina.SNP	CardioMetabo.Chip	Reference
BSND	1	55496039	rs11206510	rs11206510	rs11206510	48
CELSR2	1	109818530	rs646776	rs646776	rs646776	32
MIA3	1	222823529	rs17465637	rs17465637	rs17465637	38
						48
WDR12	2	203745885	rs6725887	rs6725887	rs6725887	48
NYAP2	2	227068080	rs2943634	rs2943634	rs2943634	38
INPP5D	2	233998481	rs10933436	rs10933436	rs10933436	48
BTD	3	15648004	rs7651039	rs7651039	rs7651039	48
MRAS	3	138122122	rs9818870	rs9818870	rs9818870	41
PHACTR1	6	12927544	rs12526453	rs12526453	rs12526453	48
ANKS1A	6	35034800	rs17609940	rs17609940	rs2077750	48
VEGFA	6	43758873	rs6905288	rs6905288	rs6905288	384
FAM46A	6	82515315	rs16893526	rs16893526	rs16893526	385
TCF21	6	134214525	rs12190287	rs12190287	rs12190287	48
MTHFD1L	6	151252985	rs6922269	rs6922269	rs6922269	38
FNDC1	6	159646333	rs365302	rs365302	rs2782552	385
LPA	6	160961137	rs3798220	rs3798220	rs3798220	50
LPA	6	161010118	rs10455872	directly typed	rs10455872	50
BCAP29	7	107244545	rs10953541	rs10953541	rs12539895	49
ASZ1	7	117067822	rs7808424	rs7808424	rs4148709	48
ZC3HC1	7	129663496	rs11556924	rs11556924	rs11556924	48
CDKN2B-AS1	9	22031005	rs7865618	rs7865618	rs7865618	385
CDKN2B-AS1	9	22125503	rs1333049	rs1333049	rs1333049	38
						385
						47
						386
ABO	9	136142203	rs514659	rs514659	rs514659	43
ABO	9	136154168	rs579459	rs579459	rs579459	48
KIAA1462	10	30316072	rs3739998	rs3739998	rs2487927	387
KIAA1462	10	30335122	rs2505083	rs2505083	rs2505083	32
LINC00619	10	44775824	rs1746048	rs1746048	rs1746048	48
LIPA	10	91002927	rs1412444	rs1412444	rs2246833	385
						32
CNNM2	10	104719096	rs12413409	rs12413409	rs12413409	48
DYNC2H1	11	103660567	rs974819	rs974819	rs974819	32
ZNF259	11	116648917	rs964184	rs964184	rs6589564	48
ST3GAL4	11	126281897	rs4937126	rs4937126	rs7933887	48
SH2B3	12	111884608	rs3184504	rs3184504	rs10774625	48
HNF1A	12	121435587	rs2259816	rs2259816	rs2259816	41
COL4A2	13	110960712	rs4773144	rs4773144	rs4773144	48

Gene	CHR	BP	SNP	Affymetrix.Illumina.SNP	CardioMetabo.Chip	Reference
HHIPL1	14	100133942	rs2895811	rs2895811	rs2895811	[48]
SMAD3	15	67458639	rs17228212	rs17228212	rs17228212	[38]
MORF4L1	15	79111093	rs4380028	rs4380028	rs4380028	[32]
SMG6	17	2125605	rs1231206	rs1231206	rs143499	[48]
SMG6	17	2126504	rs216172	rs216172	rs143499	[48]
PEMT	17	17543722	rs12936587	rs12936587	rs12936587	[48]
UBE2Z	17	46988597	rs46522	rs46522	rs15563	[48]
SMARCA4	19	11163601	rs1122608	rs1122608	rs1122608	[48]
LINC00310	21	35599128	rs9982601	rs9982601	rs9982601	[48]
SEZ6L	22	26689635	rs688034	rs688034	rs653361	[47]

Appendix 7

nature genetics

Large-scale association analysis identifies new risk loci for coronary artery disease

The CARDIoGRAMplusC4D Consortium[1]

Coronary artery disease (CAD) is the commonest cause of death. Here, we report an association analysis in 63,746 CAD cases and 130,681 controls identifying 15 loci reaching genome-wide significance, taking the number of susceptibility loci for CAD to 46, and a further 104 independent variants ($r^2 < 0.2$) strongly associated with CAD at a 5% false discovery rate (FDR). Together, these variants explain approximately 10.6% of CAD heritability. Of the 46 genome-wide significant lead SNPs, 12 show a significant association with a lipid trait, and 5 show a significant association with blood pressure, but none is significantly associated with diabetes. Network analysis with 233 candidate genes (loci at 10% FDR) generated 5 interaction networks comprising 85% of these putative genes involved in CAD. The four most significant pathways mapping to these networks are linked to lipid metabolism and inflammation, underscoring the causal role of these activities in the genetic etiology of CAD. Our study provides insights into the genetic basis of CAD and identifies key biological pathways.

Coronary artery disease and its main complication, myocardial infarction, is the leading cause of death worldwide. Although, epidemiological studies have identified many risk factors for CAD, including plasma lipid concentrations, blood pressure, smoking, diabetes and markers of inflammation, a causal role has been proven only for some (for example, low-density lipoprotein (LDL) cholesterol and blood pressure), primarily through randomized clinical trials of drug therapy directed at the risk factor[1]. Twin and family studies have documented that a significant proportion (40–50%) of susceptibility to CAD is heritable (for a review, see ref. 2). Because genotypes are not confounded by environmental exposures, genetic analysis has the potential to define which risk factors are indeed causal and to identify pathways and therapeutic targets[3,4]. To date, genome-wide association studies (GWAS) have collectively reported a total of 31 loci, associated with CAD risk at genome-wide significance ($P < 5 \times 10^{-8}$)[5–13]. However, variants at these loci explain less than 10% of the heritability of CAD. One likely reason for this is that, given the polygenic nature of complex traits and the relatively small observed effect sizes of the loci identified, many genuinely associated variants do not reach the stringent *P*-value threshold for genome-wide significance. Indeed, there is increasing evidence that the genetic architecture of common traits involves a large number of causative alleles with very small effects[14]. Addressing this will require the discovery of additional loci while leveraging large-scale genomic data to identify the molecular pathways underlying the pathogenesis of CAD. Such discovery is facilitated by building molecular networks, on the basis of DNA, RNA and protein interactions, which have nodes of known biological function that also show evidence of association with risk variants for CAD and related metabolic traits.

In the largest GWAS meta-analysis of CAD undertaken to date by the Coronary ARtery DIsease Genome-wide Replication and Meta-analysis (CARDIoGRAM) Consortium[5], which involved 22,233 cases and 64,762 controls, in addition to loci reported at genome-wide significance, a linkage disequilibrium (LD)-pruned set of 6,222 variants achieved a nominal association *P* value of less than 0.01. Here, we test these 6,222 SNPs in a meta-analysis of over 190,000 individuals, with the primary aim of identifying additional susceptibility loci for CAD. To this end, we used the Metabochip array[15], which is a custom iSELECT chip (Illumina) containing 196,725 SNPs, designed to (i) follow-up putative associations in several cardiometabolic traits, including CAD, and (ii) fine map confirmed loci for these traits. All SNPs on the array with data in the CARDIoGRAM study were considered for analysis (79,138 SNPs, of which 6,222 were the replication SNPs and 20,876 were fine-mapping SNPs in the 22 CAD susceptibility loci identified at the time at which the array was designed; the remaining SNPs were submitted by the other consortia contributing to the Metabochip array[15]). In addition, we assess whether the genome-wide significant CAD risk alleles act through traditional risk factors by considering the available large GWAS for these traits[16–20]. Finally, we identify a broader set of SNPs passing a conservative FDR threshold for association with CAD and use this set to undertake network analysis to find key biological pathways underlying the pathogenesis of CAD.

RESULTS

Study design

We expanded the CARDIoGRAM discovery data set (22,233 cases and 64,762 controls[5], stage 1) with 34 additional CAD sample collections (stage 2) of European or south Asian descent comprising 41,513 cases and 65,919 controls (study descriptions and sample characteristics are given in **Supplementary Tables 1a** and **2a**, respectively) and undertook a 2-stage meta-analysis to test SNPs on the Metabochip array

[1]A full list of authors and affiliations appears at the end of the paper.

Received 24 April; accepted 2 November; published online 2 December 2012; doi:10.1038/ng.2480

for disease association in a total of 63,746 cases and 130,681 controls. A further set of 3,630 cases and 11,983 controls from 4 independent studies was used for replication of SNPs that reached $5 \times 10^{-8} < P < 1 \times 10^{-6}$ in combined stage 1 and 2 analysis (stage 3; **Supplementary Tables 1b** and **2b**). An overview of the study design is provided in **Supplementary Figure 1**. Cases were selected for inclusion following the standard criteria for CAD and myocardial infarction used in the CARDIoGRAM study[5] (details for the stage 2 and 3 cohorts are given in **Supplementary Table 2**). Collections were typed with either the Metabochip array (60% of samples) or provided GWAS data imputed using HapMap (**Supplementary Table 3**). We applied standard quality control criteria to each study and corrected for population stratification if λ_{GC} was ≥1.05 (estimated for samples typed on the Metabochip using 4,310 SNPs associated with long QT syndrome and located at least 5 Mb away from established CAD risk loci; Online Methods). Case-control association analyses were adjusted for sex and age. For the 79,138 SNPs on the Metabochip with both stage 1 and 2 data, we combined (2-sided) *P* values from stage 1 with their respective (1-sided) *P* values for stage 2 using Fisher's method (Online Methods). In stage 3, we validated SNPs at $5 \times 10^{-8} < P < 1 \times 10^{-6}$ and combined evidence across all stages (1– 3) using a sample size–weighted meta-analysis.

Genome-wide significant loci

We first examined the 30 CAD risk loci previously reported in individuals of European ancestry at genome-wide significance (the *ADTRP* (*C6orf105*) locus has been reported only in Chinese)[12] in the stage 2 samples. For the 26 loci in which we could test the known lead SNP or a suitable proxy ($r^2 > 0.8$), we found highly significant associations in the stage 2 samples (**Table 1**). Notably, in four of these loci (*CDKN2B-AS1*, *COL4A2*, *CXCL12* and *APOE*), we detected additional SNPs not in LD ($r^2 < 0.5$) with the lead SNP, which also reached genome-wide significance and were conditionally independent when analyzed with GCTA software[21]. The additional SNP in the *APOE* locus, rs445925 ($P = 9.42 \times 10^{-11}$; $r^2 = 0.015$ with rs207560 in 1000 Genomes Project data), is located near *APOC1*, a gene previously suggested to confer risk for CAD[22]. The r^2 value between rs445925 ($P = 9.42 \times 10^{-11}$; $n = 31$ studies) and rs7412 ($P = 8.86 \times 10^{-4}$; $n = 21$ studies), which tags the *APOE* e2 allele,is 0.588. The *LIPA* locus also harbors a strong independent signal, which, however, did not reach genome-wide significance. Findings for the strongest associated variant available on the Metabochip for the other four loci (*MIA3*, 7q22, *ZNF259-APOA5-APOA1* and *ADAMTS7*) for which we did not have a good proxy for the previously reported lead SNP are also given (**Table 1**). Notably, for *ADAMTS7*, rs7173743 ($r^2 = 0.38$ with rs3825807, the published lead SNP) also achieved genome-wide significance.

We next examined the association of the 6,222 SNPs with $P < 0.01$ in CARDIoGRAM (we excluded SNPs in all loci listed in **Table 1**). Distribution of the absolute *z* scores for these SNPs in the stage 2 samples showed strong enrichment in positive scores corresponding to SNPs with directionally consistent signals between stages 1 and 2 under the null distribution, which is defined by mean = 0 and s.d. = 1 (4,260 SNPs observed versus 3,111 SNPs expected; binomial 2-sided $P = 7.5 \times 10^{-187}$) (**Supplementary Fig. 2**). In total, 19 loci showed association at $P < 1 \times 10^{-6}$ in the combined stage 1 and 2 analysis, with 13 of them reaching genome-wide significance, namely *IL6R*, *APOB*, *VAMP5-VAMP8-GGCX*, *SLC22A4-SLC22A5*, *ZEB2-AC074093.1*, *GUCY1A3*, *KCNK5*, *LPL*, *PLG*, *TRIB1*, *ABCG5-ABCG8*, *FURIN-FES* and *FLT1* (**Table 2**; Forest and regional association plots are given in **Supplementary Figs. 3** and **4**, respectively). The 6 loci with associations not reaching $P < 5 \times 10^{-8}$ were further validated (stage 3) in 4 independent studies (3,630 cases and 11,983 controls; **Supplementary Table 1b**). Two loci, *EDNRA* and *HDAC9* replicated at $P < 0.05$ and reached genome-wide significance in a combined analysis of stages 1–3 (**Table 2**); findings for those SNPs not meeting the above criteria are shown in **Supplementary Table 4**.

Of the newly associated loci reaching genome-wide significance, *TRIB1* and *ABCG5-ABCG8* were recently reported to reach study-wide significance ($P < 3 \times 10^{-6}$) in a large candidate gene (IBC array) study of CAD[13]. The same study reported rs2706399 in the *IL5* locus, which is located 200,349 bp away from the SNP we detected in the *SLC22A4-SLC22A5* locus (rs273909; **Table 2**). Although located in the same recombination interval, these SNPs are not in LD ($r^2 = 0.02$), and conditional analysis in a subset of 85,136 samples (up to19,200 cases) suggested that the 2 signals are conditionally independent; when conditioning on rs2706399 (*IL5* locus), the *P* value for rs273909 (*SLC22A4* locus) was 5.54×10^{-3} (1.33×10^{-3} initially), whereas the converse conditioning gave a *P* value of 3.34×10^{-2} for rs2706399 (*IL5*; 7.55×10^{-3} initially). We also detected a second signal in the *FES* locus (rs2521501; $P = 1.31 \times 10^{-9}$); conditional analysis with rs17514846 and rs2521501 ($r^2 = 0.43$ in 1000 Genomes Project data) showed the two signals not only to be independent but to also increase in strength upon conditioning (rs17514846 associated at $P = 1.07 \times 10^{-25}$ when conditioned on rs2521501; conversely, the *P* value for rs2521501 was 9.24×10^{-26}).

Subgroup analyses

Genetic risk of CAD could vary by age and gender and could also specifically influence the risk of its main adverse outcome, myocardial infarction[23]. We therefore undertook exploratory association analyses in subgroups partitioned by either gender, age at event (with individuals of <50 years of age being defined as young cases) or history of myocardial infarction (Online Methods). For the 46 genome-wide significant CAD risk loci, we observed no trend for higher odds ratios (ORs) in any of the subgroup analyses (**Supplementary Table 5**). However, one new locus reached genome-wide significance in males and in young CAD cases (rs16986953; $P = 1.89 \times 10^{-8}$ and 1.67×10^{-8}, respectively), which is located in a gene desert (with nearest transcript AK097927), 1.3 Mb away from the *APOB* gene. Interaction analysis conducted in a subset of studies ($n = 12$) where we had individual-level data provided suggestive evidence of an association with age ($P = 0.033$) but not with sex ($P = 0.708$); further studies are required to confirm this finding.

Wider Metabochip content

In addition to SNPs provided by the CARDIoGRAM Consortium, the Metabochip array contains a further 113,248 SNPs submitted for a range of cardiometabolic traits[15] other than CAD itself (associated at $P > 0.01$ with CAD in CARDIoGRAM samples or not tested). For these SNPs, we did not detect any new locus reaching genome-wide significance in our data set (including stage 1 and 3 data, when available). In total, therefore, we discovered 15 newly associated loci at genome-wide significance, increasing the total number of genome-wide significant loci to 45 in individuals of European and south Asian ancestry.

Localizing candidate CAD genes

To identify potential causal CAD-associated genes at the 15 new susceptibility loci identified in our study, we first analyzed genome-wide expression quantitative trait locus (eQTL) data in multiple tissues (circulating monocytes, liver, fat, skin, omentum, aortic media and adventitia, mammary artery and lymphoblastoid cell lines (LCLs)). We found that the lead SNP or a proxy in high LD ($r^2 \geq 0.8$) in three of the new loci was associated in *cis* with variable expression levels of the *GGCX-VAMP8*, *PLG* and *FES* genes (**Supplementary Table 6**).

Table 1 Association findings for known CAD susceptibility loci

Known loci[a]	Published lead SNP or proxy	New SNP (r^2 with lead SNP)	Chr.	Effect/non-effect allele (frequency)	Stage 2 OR	Stage 2 *P*	Combined *P*	Combined OR
SORT1[b]	rs602633 (tagging rs599839; $r^2 = 1.00$)		1	C/A (0.77)	1.13	2.19×10^{-18}	1.47×10^{-25}	1.12
PCSK9	rs11206510		1	T/C (0.84)	1.04	5.09×10^{-3}	1.79×10^{-5}	1.06
WDR12	rs6725887		2	C/T (0.11)	1.10	5.29×10^{-8}	1.16×10^{-15}	1.12
MRAS	rs9818870		3	T/C (0.14)	1.05	1.83×10^{-3}	2.62×10^{-9}	1.07
TCF21	rs12190287		6	C/G (0.59)	1.04	6.48×10^{-4}	4.94×10^{-13}	1.07
SLC22A3-LPAL2-LPA	rs3798220		6	C/T (0.01)	1.28	4.90×10^{-5}	N/A	N/A
		rs2048327 (0.03)	6	G/A (0.35)	1.05	1.09×10^{-5}	6.86×10^{-11}	1.06
ZC3HC1	rs11556924		7	C/T (0.65)	1.08	1.45×10^{-9}	6.74×10^{-17}	1.09
CDKN2BAS1	rs1333049		9	C/G (0.47)	1.21	1.08×10^{-34}	1.39×10^{-52}	1.23
		rs3217992 (0.50)	9	A/G (0.38)	1.14	7.27×10^{-32}	7.75×10^{-57}	1.16
ABO	rs579459		9	C/T (0.21)	1.04	2.13×10^{-2}	2.66×10^{-8}	1.07
CYP17A1-CNNM2-NT5C2	rs12413409		10	G/A (0.89)	1.08	4.12×10^{-3}	6.26×10^{-8}	1.10
KIAA1462	rs2505083		10	C/T (0.42)	1.06	2.82×10^{-7}	1.35×10^{-11}	1.06
PDGFD	rs974819		11	A/G (0.29)	1.08	2.03×10^{-9}	3.55×10^{-11}	1.07
SH2B3	rs3184504		12	T/C (0.40)	1.07	6.13×10^{-7}	5.44×10^{-11}	1.07
COL4A1-COL4A2	rs4773144		13	G/A (0.42)	1.06	2.34×10^{-6}	1.43×10^{-11}	1.07
		rs9515203 (0.01)	13	T/C (0.74)	1.08	1.13×10^{-8}	5.85×10^{-12}	1.08
HHIPL1	rs2895811		14	C/T (0.43)	1.04	1.18×10^{-4}	4.08×10^{-10}	1.06
RAI1-PEMT-RASD1	rs12936587		17	G/A (0.59)	1.04	2.06×10^{-4}	1.24×10^{-9}	1.06
LDLR	rs1122608		19	G/T (0.76)	1.06	3.72×10^{-6}	6.33×10^{-14}	1.10
Gene desert (*KCNE2*)	rs9982601		21	T/C (0.13)	1.10	8.69×10^{-9}	7.67×10^{-17}	1.13
PPAP2B	rs17114036		1	A/G (0.91)	1.09	2.68×10^{-5}	5.80×10^{-12}	1.11
ANKS1A	rs12205331 (tagging rs17609940; $r^2 = 0.85$)		6	C/T (0.81)	1.01	4.36×10^{-1}	4.18×10^{-5}	1.04
PHACTR1	rs9369640 (tagging rs12526453; $r^2 = 0.90$)		6	A/C (0.65)	1.09	1.11×10^{-12}	7.53×10^{-22}	1.09
CXCL12	rs501120		10	A/G (0.83)	1.06	7.13×10^{-5}	1.79×10^{-9}	1.07
		rs2047009 (0.05)	10	C/A (0.48)	1.05	9.66×10^{-6}	1.59×10^{-9}	1.05
LIPA	rs2246833 (tagging rs1412444; $r^2 = 0.98$)		10	T/C (0.38)	1.04	2.76×10^{-2}	9.49×10^{-6}	1.06
		rs11203042 (0.39)	10	T/C (0.44)	1.03	9.86×10^{-3}	6.08×10^{-6}	1.04
UBE2Z	rs15563 (tagging rs46522; $r^2 = 0.93$)		17	C/T (0.52)	1.01	2.44×10^{-1}	9.37×10^{-6}	1.04
SMG6	rs2281727 (tagging rs216172; $r^2 = 0.96$)		17	C/T (0.36)	1.04	8.46×10^{-4}	7.83×10^{-9}	1.05
ApoE-ApoC1	rs2075650		19	G/A (0.14)	1.11	5.86×10^{-11}	N/A	N/A
		rs445925 (0.03)	19	C/T (0.90)	1.13	8.76×10^{-9}	N/A	N/A
MIA3	N/A	rs17464857 (0.18)	1	T/G (0.87)	1.02	1.56×10^{-1}	6.06×10^{-5}	1.05
7q22	N/A	rs12539895 (0.64)	7	A/C (0.19)	1.02	4.00×10^{-2}	5.33×10^{-4}	1.08
ZNF259-APOA5-APOA1	N/A	rs9326246 (0.63)	11	C/G (0.10)	1.04	2.90×10^{-2}	1.51×10^{-7}	1.09
ADAMTS7	N/A	rs7173743 (0.38)	15	T/C (0.58)	1.06	2.46×10^{-7}	6.74×10^{-13}	1.07

Chr., chromosome.
[a]Locus *C6orf105*, which has been reported only in Chinese and has no good proxy SNP (Utah residents of Northern and Western European ancestry (CEU) or Han Chinese in Beijing, China (CHB)) on the Metabochip. The best available proxy is rs9348953 ($r^2 = 0.01$), with combined $P = 2.81 \times 10^{-3}$. [b]rs12740374, which was reported as a functional variant in this locus and has $r^2 = 0.895$ with rs599839, has combined $P = 8.25 \times 10^{-18}$ (OR = 1.135) based on the random-effects model used (*P* in stage 2 alone was 6.48×10^{-21} under the fixed-effect model).

We then assessed allele-specific expression data in monocytes, fibroblasts and LCLs and found three loci where the lead SNP was associated with an imbalance in expression of either *LPL*, *GGCX* or *FES*; *IL6R* showed some evidence of allele-specific expression in the fibroblast sample (**Supplementary Table 6**). Finally, we examined the new CAD risk loci for genes with relevant disease trait associations in mouse knockout models; six loci harbor a gene for which a mouse knockout model has a relevant cardiovascular phenotype, namely *ABCG8*, *APOB*, *GUCY1A3*, *PLG*, *LPL* and *FES* (**Supplementary Table 7**). *PLG* is adjacent to *LPA*, and, although the *PLG* risk variant rs4252120[T] was strongly associated with elevated Lp(a) lipoprotein levels ($P = 5 \times 10^{-24}$) in 3,698 PROCARDIS cases, it was associated with CAD independent of the LPA-linked variant at rs3798220. A detailed discussion of the genes in each locus is provided in the **Supplementary Note**. Of the 30 previously reported CAD susceptibility loci in individuals of European and south Asian ancestry, mouse knockout models for the candidate genes *PEMT*, *APOE*, *LDLR*, *COL4A1*, *LIPA*, *APOA1-APOA5*, *PPAP2B* and *PCSK9* also show phenotypic characteristics directly relevant to disease (**Supplementary Table 7**). In total, approximately a third of the 45 CAD loci contain a known functionally relevant candidate gene.

Overlap with traditional risk factors

We assessed both the known and new CAD susceptibility loci for overlap of associations with a number of relevant traits for which summary statistics have been made available: lipid levels (GLGC)[16], blood pressure (ICBPG)[17], diabetes (DIAGRAM)[18], glucometabolic traits (fasting insulin and fasting glucose concentrations, HOMA-B

Table 2 Additional loci showing genome-wide significant association with CAD

				Stage 1 (18,014 cases and 40,925 controls)[a]		Stage 2 (40,365 cases and 63,714 controls)		Combined (stages 1 and 2)	Stage 3 (5,055 cases and 5,617 controls)		Combined (stages 1–3)	
SNP	Chr.	Nearest gene(s)	Effect/non-effect allele (frequency)	OR	*P*	OR	*P*	*P*	OR	*P*	*P*	Biological relevance[b]
New												
rs4845625	1	*IL6R*	T/C (0.47)	1.06	4.84×10^{-5}	1.04	3.46×10^{-5}	3.55×10^{-8}	1.09	1.58×10^{-3}	3.64×10^{-10}	2
rs515135	2	*APOB*	G/A (0.83)	1.07	8.63×10^{-4}	1.08	2.17×10^{-8}	4.80×10^{-10}	1.03	4.02×10^{-1}	2.56×10^{-10}	1
rs2252641	2	*ZEB2-AC074093.1*	G/A (0.46)	1.06	1.37×10^{-5}	1.04	1.27×10^{-4}	3.66×10^{-8}	1.00	9.54×10^{-1}	5.30×10^{-8}	
rs1561198	2	*VAMP5-VAMP8-GGCX*	A/G (0.45)	1.06	7.47×10^{-5}	1.05	2.57×10^{-6}	4.48×10^{-9}	1.07	1.75×10^{-2}	1.22×10^{-10}	A,1
rs7692387	4	*GUCY1A3*	G/A (0.81)	1.08	1.04×10^{-5}	1.06	1.89×10^{-5}	4.57×10^{-9}	1.13	5.47×10^{-4}	2.65×10^{-11}	1
rs273909	5	*SLC22A4-SLC22A5*	C/T (0.14)	1.07	3.24×10^{-3}	1.09	2.00×10^{-7}	1.43×10^{-8}	1.11	2.43×10^{-2}	9.62×10^{-10}	A,1
rs10947789	6	*KCNK5*	T/C (0.76)	1.07	6.07×10^{-5}	1.06	1.22×10^{-5}	1.63×10^{-8}	1.01	7.03×10^{-1}	9.81×10^{-9}	3
rs4252120	6	*PLG*	T/C (0.73)	1.07	1.18×10^{-5}	1.06	1.82×10^{-5}	5.00×10^{-9}	1.07	9.58×10^{-2}	4.88×10^{-10}	1
rs264	8	*LPL*	G/A (0.86)	1.11	2.99×10^{-7}	1.05	7.30×10^{-4}	5.06×10^{-9}	1.06	1.60×10^{-1}	2.88×10^{-9}	1
rs9319428	13	*FLT1*	A/G (0.32)	1.06	7.88×10^{-5}	1.05	5.70×10^{-6}	1.01×10^{-8}	1.10	1.37×10^{-3}	7.32×10^{-11}	1
rs17514846	15	*FURIN-FES*	A/C (0.44)	1.07	2.37×10^{-5}	1.05	7.35×10^{-7}	4.49×10^{-10}	1.04	3.02×10^{-1}	9.33×10^{-11}	A,1
Previously reported at array-wide level of significance ($P < 3 \times 10^{-6}$)												
Rs2954029	8	*TRIB1*	A/T (0.55)	1.06	2.79×10^{-5}	1.04	7.75×10^{-5}	4.53×10^{-8}	1.05	8.56×10^{-2}	4.75×10^{-9}	4
Rs6544713	2	*ABCG5-ABCG8*	T/C (0.30)	1.06	2.22×10^{-4}	1.06	1.57×10^{-7}	8.72×10^{-10}	0.96	3.56×10^{-1}	2.12×10^{-9}	1
New (stage 3 replication)												
Rs1878406	4	*EDNRA*	T/C (0.15)	1.10	2.37×10^{-6}	1.06	3.54×10^{-3}	1.65×10^{-7}	1.09	2.01×10^{-2}	2.54×10^{-8}	1
Rs2023938	7	*HDAC9*	G/A (0.10)	1.08	6.81×10^{-4}	1.07	5.25×10^{-5}	6.49×10^{-7}	1.13	4.09×10^{-2}	4.94×10^{-8}	1

[a]Total sample sizes do not include the CHARGE sample sizes. [b]A, *cis* eQTL in LCLs; 1, mouse model available with cardiovascular phenotype; 2, mouse model has homeostatic and immune phenotypes; 3, mouse model has respiratory, nervous system, mortality, aging, growth and renal phenotypes; 4, mouse model has growth and immune phenotypes.

(homeostatic model assessment-β score) and HOMA-IR (insulin resistance); MAGIC)[19] and anthropometric traits (GIANT)[20,24]. After applying a Bonferroni correction for the 51 independent CAD-associated alleles tested (44 loci; no data available for rs16986953 and rs2521501), 12 loci showed evidence of association ($P < 1 \times 10^{-4}$) between the lead CAD risk SNP and 1 or more plasma lipid trait (total cholesterol, LDL cholesterol, high-density lipoprotein (HDL) cholesterol and triglyceride concentration) in the expected direction (the CAD risk allele was associated with higher total cholesterol, LDL cholesterol and triglyceride concentrations and lower HDL cholesterol concentration). These lead SNPs were most strongly associated with LDL cholesterol concentration at eight loci (*APOB*, *ABCG5-ABCG8*, *PCSK9*, *SORT1*, *ABO*, *LDLR*, *APOE* and *LPA*), with triglyceride concentration at two loci (*TRIB1* and the *APOA5* cluster) and with HDL cholesterol concentration at one locus (*ANKS1A*). There was near-equivalent association for triglyceride and HDL cholesterol concentrations at one locus (*LPL*). All loci except *LPA* and *ANKS1A* showed genome-wide significance for association with a lipid trait. These results underscore the importance of LDL cholesterol as a causal CAD risk factor (**Supplementary Table 8**). At the *SH2B3* locus, the CAD risk allele for rs3184504 was associated with both lower LDL cholesterol ($P = 1.73 \times 10^{-9}$) and HDL cholesterol ($P = 4.97 \times 10^{-6}$) concentration; one likely explanation is the presence of independent variants for CAD and LDL cholesterol. Two known CAD risk loci (*CYP17A1-NT5C2* and *SH2B3*) and two of the new CAD susceptibility loci (*GUCY1A3* and *FES*) have previously been associated with systolic (SBP) and diastolic (DBP) blood pressure[17]. Significant evidence for association with DBP was also observed for *ZC3HC1* (**Supplementary Table 8**). In contrast to the results for lipid concentration and blood pressure, there was no significant association of any of the loci tested with type 2 diabetes (T2D). Consistent with this observation, none of the assessed glucometabolic traits (fasting insulin and fasting glucose concentrations, HOMA-B and HOMA-IR) were related to these CAD variants (at the *ANKS1A* locus, it was not the CAD risk SNP that was associated with fasting insulin concentration and HOMA-IR). Suggestive associations ($P < 1 \times 10^{-4}$) with body mass index (BMI) and waist-hip ratio were observed in the *CYP17A1-CNNM2-NT5C2* and *RAI1-PEMT-RASD1* loci, respectively.

Additional suggestive associations

The genome-wide significance threshold, $P < 5 \times 10^{-8}$, we used is the accepted criterion for reporting individual association signals, as for each experiment it controls the error rate among common variants to less than 5%. However, SNPs showing suggestive association with a phenotype but not meeting this genome-wide threshold are likely to include additional true positive signals in well-powered studies (**Supplementary Fig. 1**). Such SNPs may also be informative in predicting CAD risk and in constructing CAD-associated biological networks. To identify such variants, we undertook an FDR analysis to assess the proportion of false positive signals in a set of (nominally) significant SNPs[25]. The Metabochip array contains both SNPs with priors in terms of association to CAD (CARDIoGRAM study $P < 0.01$) and blocks of highly correlated SNPs in fine-mapping regions. Therefore, to normalize the distribution of SNPs considered for FDR analysis, we (i) removed all SNPs in the CAD fine-mapping regions and LD-pruned ($r^2 < 0.2$) SNPs in the non CAD fine-mapping regions and (ii) adjusted the combined *P* values of all SNPs with priors in stage 1 ($P < 0.01$) using fixed-effect inverse variance–weighted meta-analysis *P* values for all other SNPs (Online Methods). In addition, we obtained 104 SNPs at an FDR threshold of 5% and LD threshold of $r^2 < 0.2$ (**Supplementary Table 9**). The median OR for CAD for these SNPs was 1.054 (interquartile range of 0.0199) per risk allele (**Supplementary Fig. 5**).

On the basis of a heritability estimate of 40% for CAD, the combination of the known and newly associated SNPs within the 45 susceptibility loci (**Tables 1** and **2**) explains approximately 6% of the additive genetic variance of CAD. The addition of the 104 SNPs from FDR analysis increased the fraction explained to 10.6% (Online Methods).

a

Canonical pathways

Biological function	*P* value
Atherosclerosis signaling	1.67×10^{-6}
LXR/RXR activation	2.14×10^{-6}
FXR/RXR activation	5.76×10^{-6}
Acute phase response signaling	4.36×10^{-5}

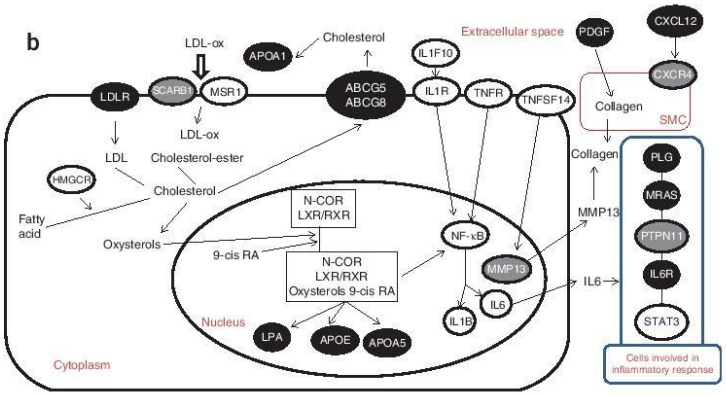

Figure 1 Canonical pathway analysis. (**a**) The four most significant canonical pathways represented in networks 3, 5 and 9, and overlapping networks ON1 (includes networks 1, 2, 6 and 8) and ON2 (includes networks 4 and 7); all molecules are listed by network in **Supplementary Table 10**. (**b**) Schematic showing parts of the atherosclerosis signaling, LXR/RXR activation and acute phase response signaling canonical pathways (Ingenuity) that are involved in both lipid metabolism and inflammation. Genes in confirmed CAD susceptibility loci (including both previously and newly reported) and in loci showing suggestive association with an FDR of <10% are depicted as black and gray ovals, respectively. Other key genes are depicted as white ovals; notably, some of them, such as *IL1F10-IL1B*, *STAT3* and *HMGCR*, have SNPs ranking in the top 1,000 in the FDR analysis. The process leading to myocardial infarction involves multiple cell types that are depicted in this schematic as a composite cell (large oval) and its nucleus (inner oval) in the extracellular space; the smooth muscle cell is shown separately (SMC; red oval), whereas the blue oval depicts cell types involved in the inflammatory response.

Network analysis

In contrast to estimating heritability where we want to keep the false positive rate as low as possible, in network analysis, we want to maximize the representation of potential network nodes in the gene set used. Thus, to perform network analysis, we selected the top 222 SNPs defined by the FDR analysis (10% FDR; final $P < 6.6 \times 10^{-4}$) at an LD threshold of $r^2 \leq 0.7$ and assigned 239 candidate genes on the basis of either eQTL data or physical proximity (**Supplementary Table 10**). We mapped 238 of the 239 genes in the Ingenuity Knowledge Base and considered 233 for network construction (Online Methods) on the basis of available data on interactions in humans, mice and/or rats (51 genes within the 46 genome-wide significant loci (set A) and 182 genes within the loci selected at FDR < 10% (set B)). Including neighboring genes, Ingenuity generated 9 networks comprising 553 nodes; these included 48 (94.1%) of the genes in set A and 156 (85.7%) of those in set B (**Supplementary Table 10**). We obtained 2 overlapping networks: ON1, which included networks 1, 2, 6 and 8, comprising the majority of genes in both sets (33 and 83 in sets A and B, respectively), and ON2, which included networks 4 and 7 (**Supplementary Table 10**). The nine networks were strongly enriched for genes (query set) known to be involved in lipid metabolism ($P = 1.48 \times 10^{-9}$), cellular movement (blood and endothelial cells; $P = 1.35 \times 10^{-7}$) and processes such as tissue morphology (size and area of atherosclerotic lesion, quantity of leukocytes, macrophages and smooth muscle cells; $P = 9.66 \times 10^{-10}$) and immune cell trafficking (migration and adhesion; $P = 1.12 \times 10^{-7}$). As a negative control in the network analysis, we used a set of 368 genes selected from the least significant SNPs in the FDR analysis; the resulting networks showed no significant enrichment in relevant molecular functions and process (results described in detail in the **Supplementary Note**).

We then assessed how genes in the networks overlap with canonical pathways in the Ingenuity database. The four most significant canonical pathways represented in these networks are shown in **Figure 1a**. The top three pathways, atherosclerosis signaling, liver X receptor (LXR)/retinoid X receptor (RXR) activation and farnesoid X receptor (FXR)/RXR activation, all harbor genes involved in lipid metabolism, including ten CAD risk loci (*ABCG5-ABCG8*, *APOA1*, *APOA5*, *APOB*, *APOE*, *CXCL12*, *LDLR*, *LPA*, *LPL* and *PDGFD*). This is in agreement with our finding that 12 CAD risk loci are associated with lipid levels at $P < 1 \times 10^{-4}$ (**Supplementary Table 8**). Notably, three of the top four pathways also contain genes involved in inflammation. In addition to the atherosclerosis signaling and LXR/RXR activation pathways, the acute phase response signaling (AAPRS) pathway, which includes four CAD risk loci (*APOA1*, *MRAS*, *IL6R* and *PLG*), is involved in inflammation and, more specifically, the rapid inflammatory response that is triggered, among other factors, by tissue injury. Genes from both the lipid metabolism and inflammation-related pathways map to all networks, except network 9, which harbors only two genes (**Supplementary Table 10**). As shown for overlapping network ON1 (**Supplementary Fig. 6**), genes in lipid metabolism and inflammation are interconnected and include both CAD-associated loci reaching genome-wide significance and candidate loci at FDR < 10%. Key interactions between CAD susceptibility genes (known, new and the FDR set) involved in lipid metabolism and inflammation are shown in **Figure 1b**; macrophages take up oxidized LDL (ox-LDL) through their cell surface scavenger receptors to form foam cells. Foam cells secrete proinflammatory cytokines, such as interleukin (IL)-1, IL-6 and matrix metalloproteinases, which can amplify the local inflammatory response and stimulate smooth muscle cell proliferation and initial migration toward the lesion[26]. Regulation of collagen secretion by smooth muscle cells in the extracellular matrix is regulated by matrix metalloproteinases. Reduction of collagen in the extracellular matrix will destabilize the plaque. Both *COL4A1* and *COL4A2* encode subunits of type IV collagen, which is the major structural component of basement membranes lining the inner surface of blood vessels. Metalloproteinases have a role in the maintenance of the extracellular matrix and remodeling, contributing to the transition of plaques from stable to vulnerable states (**Fig. 1b**).

DISCUSSION

Here, we report the largest genetic study to date assessing the impact of common variation on CAD risk. On the basis of analyses involving 63,746 CAD cases and 130,681 controls, we identified 15 new risk alleles at genome-wide significance, bringing the total number of confirmed CAD susceptibility loci in individuals of European and south Asian ancestry to 45. We also identified a further set of 104 likely independent ($r^2 < 0.2$) SNPs associated at an FDR of 5% with ORs between 1.031 and 1.126 per risk allele. In total, we estimate that these variants explain approximately 10.6% of the additive genetic variance of CAD (although we note that this may be an overestimate, given that it was not obtained in an independent sample). Our data also support the presence of additional true signals among the tested common SNPs that are likely to further contribute in explaining heritability; for example, the *P*-value adjustment we applied in the FDR analysis penalized the replication SNPs.

Among the 45 loci in individuals of European and south Asian ancestry that were confirmed to be associated with CAD, we found that 12 were significantly associated with the concentrations of blood lipids (mainly with LDL cholesterol), and 5 were associated with blood pressure. These data support the known etiological relationships of plasma lipids and blood pressure with CAD. People with T2D seem to have a 1.5- to 2-fold higher risk of CAD than those without diabetes[27], but none of the 45 risk loci were associated with diabetes status or with continuous levels of various glucometabolic traits. We note that, for the binary variable of T2D status, inability to show associations with CAD risk loci may reflect limited statistical power. The temporal relationship for comorbidity with both diabetes and CAD is complex: individuals with CAD without diabetes at diagnosis often subsequently develop T2D[28]. Furthermore, despite clear benefits in preventing microvascular disease (for example, retinopathy and nephropathy), intensive glucose control in diabetics reduces the risk of cardiovascular disease relatively modestly[29]. However, before a final conclusion can be reached, as many cohorts contributing to this meta-analysis focused by design on early disease manifestation or excluded diabetic individuals, a formal testing of the relationship of T2D and CAD in Mendelian randomization experiments will be necessary. To this end, the large genetic association data set on CAD assembled here will also facilitate testing of the causal relationship of other putative risk factors for CAD.

A desirable clinical goal is to integrate genetic information into a risk score for CAD in an attempt to provide improved predictive power over traditional risk factors in asymptomatic subjects, such that preventative measures, where available, can be more appropriately targeted. Our findings provide an appropriate framework of 153 CAD risk variants (at those established as susceptibility loci meeting the genome-wide significance threshold and additional suggestive loci with an FDR of <5%) for assessing a genetic risk score in well-powered prospective studies to determine whether they are sufficiently informative and independent predictors to have potential for use in day-to-day practice.

Allowing for inherent limitations in selecting likely candidate genes at each locus, our network analysis identified lipid metabolism and inflammation as key biological pathways involved in the genetic pathogenesis of CAD. Indeed, there was significant crosstalk between the lipid metabolism and inflammation pathways identified (**Fig. 1**). The emergence of lipid metabolism as a key pathway provides a positive control for the network and pathway analysis. On the other hand, this analysis provides strong new evidence at the molecular level in support of the causal involvement of inflammatory mechanisms in the pathogenesis of coronary atherosclerosis[30]. The role of inflammation in atherosclerosis is well documented in the literature[26]; for example, risk factors such as fat diet, smoking, hypertension, hyperglycemia, obesity and insulin resistance can trigger the expression of adhesion molecules (upregulated by atherogenic lipoproteins such as ox-LDL, very-low-density lipoprotein (VLDL) and Lp(a) lipoprotein) by endothelial cells, leading to the attachment of monocytes to the arterial wall. Although our analysis identified as significant the rapid inflammatory response pathway (mediated by NF-κB, MAPK and JAK-STAT signaling) that is primarily involved in innate immunity, many of the effector pathways in innate and adaptive immunity are heavily overlapping, and both are likely to have a role in CAD pathogenesis[26]. The five CAD-related networks constitute a useful framework for further functional and mechanistic studies to elucidate the biological processes underlying CAD pathogenesis and to investigate gene-environment interactions.

URLs. QVALUE software for FDR analysis, http://genomics.princeton.edu/storeylab/qvalue/; coronary heart disease statistics, http://www.bhf.org.uk/publications/view-publication.aspx?ps=1002097; top 10 causes of death fact sheet 310, http://www.who.int/mediacentre/factsheets/fs310/en/index.html; Uppsala Platform, http://molmed.medsci.uu.se/SNP+SEQ+Technology+Platform/Genotyping.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Summary statistics for the 79,138 SNPs considered in this study for association with CAD (SNPs with stage 1 and stage 2 data) are available at ftp://ftp.sanger.ac.uk/pub/cardiogramplusc4d/.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the personnel of the Wellcome Trust Sanger Institute (WTSI) Genotyping Facility, in particular S. Edkins, for supervising the genotyping of the AMC-PAS, Cardiogenics, GLACIER, MORGAM, PROMIS, THISEAS, and WTCCC cohorts.

AMC-PAS/SANQUIN.
We thank A.A. Soussan for technical assistance.

We thank personnel from the Estonian Genome Center of the University of Tartu (EGCUT) and the Estonian Biocentre, especially M. Hass and V. Soo, for data generation.

FINCAVAS.
We thank the staff of the Department of Clinical Physiology for collecting the exercise test data.

The GLACIER Study.
The GLACIER study is a nested study within the Northern Sweden Health and Disease Study; phenotyping was conducted as part of the Västerbotten Intervention Project. We thank the participants and the investigators from these studies for their valuable contributions, with specific thanks to L. Weinehall, Å. Agren, K. Enquist and T. Johansson.

GoDARTS Dundee.
We are grateful to all the participants who took part in this study, to the general practitioners, to the Scottish School of Primary Care for their help in recruiting the participants and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We acknowledge the support of the Health Informatics Centre at the University of Dundee in managing and supplying the anonymized data and National Health Service (NHS) Tayside, the original data owner.

Heart Protection Study.
The study was designed and conducted by the Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU) at the University of Oxford. Genotyping was supported by a grant to Oxford University and Centre National de Genotypage (CNG) from Merck. The funders had no role in the design of the study or in the data collection or analysis. We especially acknowledge the participants in the study,

the Steering Committee and our collaborators. J.C.H. acknowledges support from the British Heart Foundation (BHF) Centre of Research Excellence.

LOLIPOP.
We thank the participants and research staff who made the study possible.

MORGAM study.
We thank the contributing sites and key personnel, as detailed below.

Finland: We thank FINRISK, National Institute for Health and Welfare, Helsinki: V.S. (principal investigator), A. Juolevi, E. Vartiainen and P. Jousilahti; Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) study, National Institute for Health and Welfare, Helsinki: J. Virtamo (principal investigator) and H. Kilpeläinen; the MORGAM Data Centre, National Institute for Health and Welfare, Helsinki: K. Kuulasmaa (responsible person), Z. Cepaitis, A. Haukijärvi, B. Joseph, J. Karvanen, S. Kulathinal, M. Niemelä and O. Saarela; and the MORGAM Central Laboratory, National Institute for Health and Welfare, Helsinki: M.P. (responsible person), P. Laiho and M. Sauramo.

France: We thank the National Coordinating Centre, National Institute of Health and Medical Research (U258), Paris: P. Ducimetière (national coordinator) and A. Bingham; Prospective Epidemiological Study of Myocardial Infarction (PRIME)/Strasbourg, Department of Epidemiology and Public Health, EA 3430, University of Strasbourg, Faculty of Medicine, Strasbourg: D. Arveiler (principal investigator), B. Haas and A. Wagner; PRIME/Toulouse, Department of Epidemiology, Toulouse University School of Medicine, Toulouse: J.F. (principal investigator), J.-B. Ruidavets, V. Bongard, D. Deckers, C. Saulet and S. Barrere; PRIME/Lille, Department of Epidemiology and Public Health, INSERM U744–Université Lille Nord de France–Institut Pasteur de Lille, Lille: P. Amouyel (principal investigator), M. Montaye, B. Lemaire, S. Beauchant, D. Cottel, C. Graux, N. Marecaux, C. Steclebout and S. Szeremeta; and the MORGAM Laboratory, INSERM U937, Paris: F.C. (responsible person), L. Tiret and V. Nicaud.

Italy: We thank Centro Ricerche EPIMED–Epidemiologia e Medicina Preventiva, Dipartimento di Medicina Clinica e Sperimentale; Università dell' Insubria, Varese: M.M.F. (principal investigator) and G. Veronesi; and Research Centre on Public Health, University of Milano–Bicocca, Monza: G. Cesana.

UK: We thank PRIME/Belfast, Queen's University Belfast, Belfast: F.K. (principal investigator), A.E. (former principal investigator), J. Yarnell and E. Gardner; and the MORGAM Coordinating Centre, Queen's University Belfast, Belfast: A.E. (MORGAM coordinator), S. Cashman and F.K.
MORGAM management group: A.E. (chair), S.S.B., F.C., M.M.F., K. Kuulasmaa, A. Palotie, M.P., A.P., V.S., H. Tunstall-Pedoe and P.G. Wiklund. Previous members: K. Asplund, L. Peltonen, D. Shields and B. Stegmayr. The PRIME Study is organized under an agreement between INSERM and the Merck, Sharpe and Dohme-Chibret Laboratory, with the following participating laboratories: The Strasbourg MONICA Project, Laboratoire d'Epidémiologie et de Santé Publique, and the Université de Strasbourg, Strasbourg, France (D. Arveiler and B. Haas); The Toulouse MONICA Project, UMR INSERM 1027, and the Department of Epidemiology, Toulouse University School of Medicine, Université Paul Sabatier, Toulouse, France (J.F. and J.-B. Ruidavets); The Lille MONICA Project, INSERM U744, Institut Pasteur de Lille and Université Lille Nord de France, Lille, France (P. Amouyel and M. Montaye); The Department of Epidemiology and Public Health, Queen's University, Belfast, Belfast, UK (A.E., J. Yarnell and F.K.); The Department of Atherosclerosis, INSERM U545, Institut Pasteur de Lille, Faculté de Médecine and Université Lille Nord de France, Lille, France (G. Luc and J.-M. Bard); The Laboratory of Haematology, INSERM U626, and Hôpital La Timone, Marseille, France (I. Juhan-Vague and P. Morange); The Laboratory of Endocrinology, INSERM U563, Toulouse, France (B. Perret); The Vitamin Research Unit, The University of Bern, Bern, Switzerland (F. Gey); The Nutrition and Metabolism Group, Centre for Public Health, Queen's University Belfast, Belfast, UK (J. Woodside and I. Young); The DNA Bank, INSERM/Université Pierre et Marie Curie (UPMC), Paris Université Unite Mixte de Recherche (UMRS) 937, Paris (F.C.); The Coordinating Centre, Institut Fédératif de Recherche Santé Publique (IFR 69), Villejuif, France (P. Ducimetière); and INSERM U970, Villejuif, France, and University Paris V, Paris Cardiovascular Research Centre (PAARC), Paris (A. Bingham).

PIVUS/Swedish Twin Registry.
We thank the SNP&SEQ Technology Platform in Uppsala (see URLs) for genotyping, in particular T. Axelsson, A.-C. Wiman and C. Pöntinen for excellent assistance.

Ulm (EMIL).
We thank the Centre of Excellence Baden-Wuerttemberg Metabolic Disorders.

WTCCC.
We thank the BHF Family Heart Study Research Group for the collection of the cases.

AUTHOR CONTRIBUTIONS

Writing committee: P. Deloukas, S. Kanoni, C.W., M.F., T.L.A., J.R.T., E.I., D. Saleheen, J.E., M.P. Reilly, R. Collins, S. Kathiresan, A.H., U.T., J.S.K., J.D., C.N.A.P., R.R., H.W., H.S. and N.J.S. **Steering committee:** P. Deloukas, S. Kanoni, C.W., M.F., T.L.A., J.R.T., E.I., D. Saleheen, J.E., M.P. Reilly, R. Collins, S. Kathiresan, A.H., U.T., J.S.K., J.D., C.N.A.P., R.R., H.W., H.S., N.J.S., S.S.B., B.O.B., J.C.C., R. Clarke, G.D., P.W.F., C.H., G.K.H., Jong-Young Lee, T.L., W.M., A.M., M.S.N., C.O., M.P., S. Ripatti, M.S.S., S.S., A. Siegbahn, C.J.W. and P.A.Z. **Analysis committee:** B.A.G., K. Stirrups, I.R.K., J.-B.C., Å.J., T.E., L.F., A.G., A.S. Havulinna, W.K.H., J.C.H., N.E., M.E.K., K. Kristiansson, P.L., L.-P.L., S. Rafelt, D. Shungin, R.J.S., G. Thorleifsson, E.T., N.V.Z., B.F.V., L.L.W., W.Z. and A.Z. **Genotyping:** D. Absher, I.B., C.B., S.C.-B., DIAGRAM Consortium, N.E.M., K.F., P.F., B.G., L.G., S.G., J.H., B.-G.H., S.E.H., T.K., J.W.K., C. Langenberg, C. Langford, M.I.M., M.M.-N., K.N., J.F.P., S. Rosinger, D.R., M.P. Rumpf, A. Schäfer, A.F.R.S., P.J.W. and Wellcome Trust Case Control Consortium. **Array design:** H.M.K. and N.W.R. **Functional analyses:** E.G., P.E., A.F.-C., A.L., O.M., S.M., MuTHER Consortium, T.-P.Y., A.H.G., E.S., T.P. and A.-C.S. **Samples and phenotyping:** (ADVANCE) A.S.G., C.I. and T.Q.; (AMC-PAS/SANQUIN) C.E.v.d.S. and H.B.; (Angio-Lüb/KORA) P. Diemert; (CADomics) P.S.W.; (CARDIOGENICS) F.C. and W.H.O.; (CHARGE) E.B., A.L.C., A.D. and V.G.; (Corogene) M.-L.L. and J.S.; (deCODE) G. Thorgeirsson, H.H. and K. Stefansson; (EPIC-NORFOLK) N.W.; (Estonian Biobank) E.M.; (FGENTCARD) D.G.; (FINCAVAS) M.K.; (FINRISK 2007/DILGOM) V.S.; (FRISCII) L.W.; (GerMIFS) T.I., C.M., K. Stark and M.E.Z.; (GLACIER) G.H.; (GoDARTS Dundee) A.S.F.D. and A.D.M.; (HPS) S.P.; (Korean GenRIC) Y.J., H.-S.K., Ji-Young Lee and J.E.P.; (LOLIPOP) S.-T.T.; (LURIC/AtheroRemo) R.L. and W. Koenig; (METSIM) J.K., M.B. and M.L.; (MIGen) R.D.; (MORGAM) K. Kuulasmaa, J.V., P.A., D. Arveiler., J.F., D.-A.T., N.K., A.P., P.B., M.M.F., A.E. and F.K.; (Ottawa Heart Genomics Study) G.A.W., S.L.H. and S.H.S.; (PennCATH/MedStar) S.E.E. and D.J.R.; (Pfizer-Broad-Malmo) D. Altshuler and D.C.; (PIVUS/Swedish Twin Registry) C.S., L.L. and N.L.P.; (PROMIS) A.R.; (SHEEP-SCARF) K.L. and U.d.F.; (THISEAS) M.D., G.K.; (Ulm-EMIL) W. Kratzer; and (WTCCC) A.J.B., P.S.B., M.S. and A.S. Hall.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2480.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Berry, J.D. *et al.* Lifetime risks of cardiovascular disease. *N. Engl. J. Med.* **366**, 321–329 (2012).
2. Peden, J.F. & Farrall, M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum. Mol. Genet.* **20** R2, R198–R205 (2011).
3. Kingsmore, S.F., Lindquist, I.E., Mudge, J., Gessler, D.D. & Beavis, W.D. Genome-wide association studies: progress and potential for drug discovery and development. *Nat. Rev. Drug Discov.* **7**, 221–230 (2008).
4. Stein, E.A. *et al.* Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. *N. Engl. J. Med.* **366**, 1108–1118 (2012).
5. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
6. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43**, 339–344 (2011).
7. Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* **361**, 2518–2528 (2009).
8. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
9. Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.* **41**, 280–282 (2009).
10. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
11. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41**, 1182–1190 (2009).
12. Wang, F. *et al.* Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat. Genet.* **43**, 345–349 (2011).
13. The IBC 50K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* **7**, e1002260 (2011).
14. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
15. Voight, B.F. *et al.* The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
16. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

17. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
18. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
19. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
20. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
21. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
22. Ken-Dror, G., Talmud, P.J., Humphries, S.E. & Dreno, F. *APOE/C1/C4/C2* gene cluster genotypes, haplotypes and lipid levels in prospective coronary heart disease risk among UK healthy men. *Mol. Med.* **16**, 389–399 (2010).
23. Reilly, M.P. *et al.* Identification of *ADAMTS7* as a novel locus for coronary atherosclerosis and association of *ABO* with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
24. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
25. Storey, J.D. & Tibshirani, R. Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol. Biol.* **224**, 149–157 (2003).
26. Hansson, G.K., Libby, P., Schönbeck, U. & Yan, Z.-Q. Innate and adaptive immunity in the pathogenesis of atherosclerosis. *Circ. Res.* **91**, 281–291 (2002).
27. Emerging Risk Factors Collaboration. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N. Engl. J. Med.* **364**, 829–841 (2011).
28. Mozaffarian, D. *et al.* Incidence of new-onset diabetes and impaired fasting glucose in patients with recent myocardial infarction and the effect of clinical and lifestyle risk factors. *Lancet* **370**, 667–675 (2007).
29. Turnbull, F.M. *et al.* Intensive glucose control and macrovascular outcomes in type 2 diabetes. *Diabetologia* **52**, 2288–2298 (2009).
30. Ross, R. Atherosclerosis is an inflammatory disease. *Am. Heart J.* **138**, S419–S420 (1999).

The authors of this paper are:
Panos Deloukas[1,126], Stavroula Kanoni[1,126], Christina Willenborg[2,126], Martin Farrall[3,4,126], Themistocles L Assimes[5,126], John R Thompson[6,126], Erik Ingelsson[7,126], Danish Saleheen[8–10,126], Jeanette Erdmann[2,126], Benjamin A Goldstein[5], Kathleen Stirrups[1], Inke R König[11], Jean-Baptiste Cazier[4], Åsa Johansson[12], Alistair S Hall[13], Jong-Young Lee[14], Cristen J Willer[15,16], John C Chambers[17], Tõnu Esko[18,19], Lasse Folkersen[20,21], Anuj Goel[3,4], Elin Grundberg[22], Aki S Havulinna[23], Weang K Ho[10], Jemma C Hopewell[24,25], Niclas Eriksson[12], Marcus E Kleber[26,27], Kati Kristiansson[23], Per Lundmark[28], Leo-Pekka Lyytikäinen[29,30], Suzanne Rafelt[31], Dmitry Shungin[32–34], Rona J Strawbridge[20,21], Gudmar Thorleifsson[35], Emmi Tikkanen[36,37], Natalie Van Zuydam[38], Benjamin F Voight[39], Lindsay L Waite[40], Weihua Zhang[17], Andreas Ziegler[11], Devin Absher[40], David Altshuler[41–44], Anthony J Balmforth[45], Inês Barroso[1,46], Peter S Braund[31,47], Christof Burgdorf[48], Simone Claudi-Boehm[49], David Cox[50], Maria Dimitriou[51], Ron Do[41,43], DIAGRAM Consortium[52], CARDIOGENICS Consortium[52], Alex S F Doney[38], NourEddine El Mokhtari[53], Per Eriksson[20,21], Krista Fischer[18], Pierre Fontanillas[41], Anders Franco-Cereceda[54], Bruna Gigante[55], Leif Groop[56], Stefan Gustafsson[7], Jörg Hager[57], Göran Hallmans[58], Bok-Ghee Han[14], Sarah E Hunt[1], Hyun M Kang[59], Thomas Illig[60], Thorsten Kessler[48], Joshua W Knowles[5], Genovefa Kolovou[61], Johanna Kuusisto[62], Claudia Langenberg[63], Cordelia Langford[1], Karin Leander[55], Marja-Liisa Lokki[64], Anders Lundmark[28], Mark I McCarthy[3,65,66], Christa Meisinger[67], Olle Melander[56], Evelin Mihailov[19], Seraya Maouche[68], Andrew D Morris[38], Martina Müller-Nurasyid[69–72], MuTHER Consortium[52], Kjell Nikus[73], John F Peden[3], N William Rayner[3], Asif Rasheed[9], Silke Rosinger[74], Diana Rubin[53], Moritz P Rumpf[48], Arne Schäfer[75], Mohan Sivananthan[76,77], Ci Song[7], Alexandre F R Stewart[78,79], Sian-Tsung Tan[80], Gudmundur Thorgeirsson[81,82], C Ellen van der Schoot[83], Peter J Wagner[36,37], Wellcome Trust Case Control Consortium[52], George A Wells[78,79], Philipp S Wild[84,85], Tsun-Po Yang[1], Philippe Amouyel[86], Dominique Arveiler[87], Hanneke Basart[88], Michael Boehnke[59], Eric Boerwinkle[89], Paolo Brambilla[90], Francois Cambien[68], Adrienne L Cupples[91,92], Ulf de Faire[55], Abbas Dehghan[93], Patrick Diemert[94], Stephen E Epstein[95], Alun Evans[96], Marco M Ferrario[97], Jean Ferrières[98], Dominique Gauguier[3,99], Alan S Go[100], Alison H Goodall[31,47], Villi Gudnason[81,101], Stanley L Hazen[102], Hilma Holm[35], Carlos Iribarren[100], Yangsoo Jang[103], Mika Kähönen[104], Frank Kee[105], Hyo-Soo Kim[106], Norman Klopp[60], Wolfgang Koenig[107], Wolfgang Kratzer[108], Kari Kuulasmaa[23], Markku Laakso[62], Reijo Laaksonen[108], Ji-Young Lee[14], Lars Lind[28], Willem H Ouwehand[1,109,110], Sarah Parish[24,25], Jeong E Park[111], Nancy L Pedersen[7], Annette Peters[67,112], Thomas Quertermous[5], Daniel J Rader[113], Veikko Salomaa[23], Eric Schadt[114], Svati H Shah[115,116], Juha Sinisalo[117], Klaus Stark[118], Kari Stefansson[35,81], David-Alexandre Trégouët[68], Jarmo Virtamo[23], Lars Wallentin[12], Nicholas Wareham[63], Martina E Zimmermann[118], Markku S Nieminen[117], Christian Hengstenberg[118], Manjinder S Sandhu[1,63], Tomi Pastinen[119], Ann-Christine Syvänen[28], G Kees Hovingh[88], George Dedoussis[51], Paul W Franks[32–34,120], Terho Lehtimäki[29,30], Andres Metspalu[18,19], Pierre A Zalloua[121], Agneta Siegbahn[12], Stefan Schreiber[75], Samuli Ripatti[1,37], Stefan S Blankenberg[94], Markus Perola[23], Robert Clarke[24,25], Bernhard O Boehm[74], Christopher O'Donnell[93], Muredach P Reilly[122,126], Winfried März[26,123], Rory Collins[24,25,126], Sekar Kathiresan[41,124,125,126], Anders Hamsten[20,21,126], Jaspal S Kooner[80,126], Unnur Thorsteinsdottir[35,81,126], John Danesh[9,126], Colin N A Palmer[38,126], Robert Roberts[78,79,126], Hugh Watkins[3,4,126], Heribert Schunkert[48,126] & Nilesh J Samani[31,47,126]

[1]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [2]Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, Lübeck, Germany. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [4]Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [5]Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. [6]Department of Health Sciences, University of Leicester, Leicester, UK. [7]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [8]Center for Non-Communicable Diseases, Karachi, Pakistan. [9]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [10]Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [11]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany. [12]Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. [13]Division of Cardiovascular and Neuronal Remodelling, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK. [14]Center for Genome Science, Korea National Institute of Health, Korea Center for Disease Control and Prevention, Yeonje-ri, Chungwon-gun, Korea. [15]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA. [16]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. [17]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [18]Estonian Genome Center, University of Tartu, Tartu, Estonia. [19]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [20]Atherosclerosis Research Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden. [21]Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden. [22]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [23]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [24]Clinical Trial Service Unit, University of Oxford, Oxford, UK. [25]Epidemiological Studies Unit, University of Oxford, Oxford, UK. [26]Mannheim Institute of Public Health, Social and Preventive Medicine, Medical Faculty of Mannheim, University of Heidelberg, Mannheim, Germany. [27]Ludwigshafen Risk and Cardiovascular Health (LURIC) Study, Freiburg, Germany. [28]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [29]Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital, Tampere, Finland. [30]Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland. [31]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. [32]Genetic & Molecular Epidemiology Unit, Department of Clinical Sciences, Lund University Diabetes Center, Skåne University Hospital, Malmö, Sweden. [33]Department of Public Health & Clinical Medicine, Genetic Epidemiology & Clinical Research Group, Section for Medicine, Umeå University, Umeå, Sweden. [34]Department of Odontology, Umeå University, Umeå, Sweden. [35]deCODE Genetics, Reykjavik, Iceland. [36]Institute for Molecular Medicine FIMM, University of Helsinki, Helsinki, Finland. [37]Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland. [38]Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK. [39]Department of Pharmacology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [40]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. [41]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [42]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [43]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [44]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [45]Division of Cardiovascular and Diabetes Research, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK. [46]University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [47]National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. [48]Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. [49]Practice of Gynecology, Ulm University Medical Centre, Ulm, Germany. [50]Biotherapeutics and Bioinnovation Center, Pfizer, South San Francisco, California, USA. [51]Department of Dietetics–Nutrition, Harokopio University, Athens, Greece. [52]A list of members and affiliations appears in the **Supplementary Note**. [53]Klinik für Innere Medizin, Kreiskrankenhaus Rendsburg, Rendsburg, Germany. [54]Cardiothoracic Surgery Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [55]Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [56]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden. [57]Commissariat à l'Energie Atomique (CEA)–Genomics Institute, National Genotyping Centre, Paris, France. [58]Department of Public Health & Clinical Medicine, Section for Nutritional Research, Umeå University, Umeå, Sweden. [59]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. [60]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany. [61]First Cardiology Department, Onassis Cardiac Surgery Center 356, Athens, Greece. [62]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [63]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [64]Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland. [65]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [66]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK. [67]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [68]Institut National de la Santé et la Recherche Médicale (INSERM) Unité Mixte de Recherche (UMR) S937, Institute for Cardiometabolism and Nutrition (ICAN), Pierre and Marie Curie (Paris 6) University, Paris, France. [69]Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany. [70]Chair of Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [71]Chair of Genetic Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [72]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [73]Heart Centre, Department of Cardiology, Tampere University Hospital, Tampere, Finland. [74]Division of Endocrinology and Diabetes, Department of Internal Medicine, Ulm University Medical Centre, Ulm, Germany. [75]Institut für Klinische Molekularbiologie, Christian-Albrechts Universität, Kiel, Germany. [76]Division of Epidemiology, Multidisciplinary Cardiovascular Research Centre (MCRC) University of Leeds, Leeds, UK. [77]Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK. [78]University of Ottawa Heart Institute, Cardiovascular Research Methods Centre Ontario, Ottawa, Ontario, Canada. [79]Ruddy Canadian Cardiovascular Genetics Centre, Ottawa, Ontario, Canada. [80]National Heart and Lung Institute (NHLI), Imperial College London, Hammersmith Hospital, London, UK. [81]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [82]Department of Medicine, Landspitali University Hospital, Reykjavik, Iceland. [83]Department of Experimental Immunohematology, Sanquin, Amsterdam, The Netherlands. [84]Center for Thrombosis and Hemostasis, University Medical Center Mainz, Johannes Gutenberg University Mainz, Mainz, Germany. [85]Department of Medicine 2, University Medical Center Mainz, Johannes Gutenberg University Mainz, Mainz, Germany. [86]Institut Pasteur de Lille, INSERM U744, Université Lille Nord de France, Lille, France. [87]Department of Epidemiology and Public Health, EA3430, University of Strasbourg, Strasbourg, France. [88]Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands. [89]Human Genetics Center, University of Texas Health Science Center, Houston, Texas, USA. [90]Department of Experimental Medicine, University of Milano–Bicocca, Monza, Italy. [91]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [92]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [93]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [94]Clinic for General and Interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany. [95]Cardiovascular Research Institute, Washington Hospital Center, Washington, DC, USA. [96]Centre for Public Health, The Queen's University of Belfast, Belfast, UK. [97]Research Centre for Epidemiology and Preventive Medicine (EPIMED), Department of Clinical and Experimental Medicine, University of Insubria, Varese, Italy. [98]Department of Cardiology, Toulouse University School of Medicine, Rangueil Hospital, Toulouse, France. [99]INSERM UMR S872, Cordeliers Research Centre, Paris, France. [100]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA. [101]Icelandic Heart Association, Kopavogur, Iceland. [102]Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA. [103]Cardiology Division, Department of Internal Medicine, Cardiovascular Genome Center, Yonsei University, Seoul, Korea. [104]Department of Clinical Physiology, Tampere University Hospital and University of Tampere, Tampere, Finland. [105]UK Clinical Research Collaboration (UKCRC) Centre of Excellence for Public Health (Northern Ireland), Queen's University of Belfast, Belfast, UK. [106]Department of Internal Medicine, Cardiovascular Center, Seoul National University Hospital, Seoul, Korea. [107]Department of Internal Medicine II–Cardiology, Ulm University Medical Center, Ulm, Germany. [108]Science Center, Tampere University Hospital, Tampere, Finland. [109]Department of Haematology, University of Cambridge, Cambridge, UK. [110]National Health Service (NHS) Blood and Transplant, Cambridge, UK. [111]Division of Cardiology, Samsung Medical Center, Seoul, Korea. [112]Munich Heart Alliance, Munich, Germany. [113]Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [114]Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA. [115]Center for Human Genetics, Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA. [116]Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA. [117]Division of Cardiology, Department of Medicine, Helsinki University Central Hospital (HUCH), Helsinki, Finland. [118]Klinik und Poliklinik für Innere Medizin II, Regensburg, Germany. [119]Department of Human Genetics, McGill University, Montréal, Québec, Canada. [120]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. [121]Lebanese American University, Chouran, Beirut, Lebanon. [122]Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [123]Synlab Academy, Mannheim, Germany. [124]Cardiology Division, Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [125]Cardiovascular Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [126]These authors contributed equally to this work. Correspondence should be addressed to P. Deloukas (panos@sanger.ac.uk) or N.J.S. (njs@leicester.ac.uk).

ONLINE METHODS

Meta-analysis and combination of evidence across stages. Analyses were performed in each study (**Supplementary Table 1a**) to test the following comparisons: all CAD cases with all controls, adjusted for sex and age; male CAD cases with male controls, adjusted for age; female CAD cases with female controls, adjusted for age; early-onset CAD cases with early age of onset (≤50 years) with all CAD controls, adjusted for sex; late-onset CAD cases (>50 years) with all controls, adjusted for sex; and all myocardial infarction cases with all controls, adjusted for age and sex. Age was defined as the recruitment age for controls and the event age for cases. We used the additive genetic model and fixed-effect inverse variance–weighted meta-analysis. SNPs were excluded from the meta-analysis if present in <17 GWAS and/or Metabochip or <13 Metabochip stage 2 studies. Heterogeneity was evaluated using the Cochran's Q and I^2 statistics. For SNPs with non-significant heterogeneity (P for $Q > 0.01$), we report fixed-effect model results. For SNPs with significant heterogeneity (P for $Q < 0.01$), we performed an outlier test comparing the results in each study with the average of all other studies. For outliers ($P < 0.01$ or no studies with data), we excluded the most extreme study and repeated the meta-analysis. If no outliers were detected, but heterogeneity was significant, we used a random-effects model that was also used for all SNPs with significant heterogeneity in stage 1. In stage 3, we used a fixed-effect inverse variance–weighted meta-analysis.

The combination of evidence across stage 1 and stage 2 meta-analysis results was performed using Fisher's combined P-values method; using two-sided P values from stage 1 and one-sided P values from stage 2 for all SNPs with consistent direction of effect across the two stages. We estimated the stage 1 and 2 combined effect sizes for SNPs in the known loci using a fixed-effect inverse variance–weighted meta-analysis. The combination of evidence across stages 1– 3 for the replication effort was performed using a sample size–weighted meta-analysis for the selected SNPs. All participants gave written consent for participation in genetic studies, and the protocol of each study was approved by the corresponding local research ethics committee or institutional review board.

False discovery rate. FDR control is an alternative approach to experiment-wise error rate control that allows for statistical multiple testing; identifying as many significantly associated SNPs as possible with a tolerable false positive burden. FDR analysis is useful for selecting extended panels of SNPs (and genes) for experiments on the basis of multiple signals (for example, pathway or network analyses) that are robust to contamination by a small number of false positive signals. However, given the specific design of the Metabochip array to include selected SNPs (for replication) with significant P values and several high-density regions (for fine mapping) associated with CAD and the other cardiometabolic traits, the number of SNPs significantly associated with CAD, as well as high LD, could bias the FDR analysis. Therefore, we excluded from the 79,138 SNPs with available stage 1 and 2 data all SNPs falling in a high-density region associated with CAD (**Tables 1** and **2**), as well as CAD risk SNPs associated at $P < 5 \times 10^{-8}$. Furthermore, we performed an LD-based SNP pruning of the remaining high-density regions ($r^2 < 0.2$). In total, 54,806 SNPs were included in the FDR analysis.

We combined stage 1 and 2 data as an inverse variance–weighted average, and P values were calculated by Wald test. SNPs selected because their stage 1 P values were below 0.01 had their combined P values adjusted. If p_0 is the P value used as the criterion for selection in stage 1, z_{12} is the standardized test statistic obtained by combining the stages (arbitrarily assumed to be positive) and s_1 and s_2 are the standard errors for the two stages, then the adjusted P value is the sum of two integrals representing the two tails in which the stage 1 result might fall. The first is:

$$I_1 = \int_{z_{12}}^{\infty} \int_{\Phi^{-1}(1-\frac{p_0}{2})}^{\infty} \frac{\Phi(u)\Phi\left(s_2\left[v\sqrt{\frac{1}{s_1^2}+\frac{1}{s_2^2}}-\frac{u}{s_1}\right]\right)}{p_0} s_2\sqrt{\frac{1}{s_1^2}+\frac{1}{s_2^2}}\partial u \partial v$$

and the second has the same form but is integrated from $-\infty$ to $\Phi^{-1}(p_0/2)$, where Φ is the cumulative normal function. To test the adjusted P values, a simulation was performed in which null SNPs were generated and selected in stage 1 on the basis of their P values. These were combined with random second-stage data simulated again assuming a null effect. The adjusted P values had the expected uniform distribution between zero and one, suitable for use in the FDR analysis.

FDR analysis was performed using QVALUE software. A natural cubic spline (with 4 degrees of freedom) was fitted to provide a smoothed estimate of the proportion of null P values ($\hat{\pi}_0$). A density histogram of the P values for the 54,806 SNPs is shown in **Supplementary Figure 7**. At FDR = 0.05, we obtained 138 SNPs that were combined with 73 independent SNPs from fine-mapping regions associated with CAD. The selection included the SNP with the lowest combined P value per fine-mapping region and all SNPs within these regions that met the 5% FDR criterion in a separate analysis and were unlinked ($r^2 < 0.2$). Finally, all SNPs reported in **Tables 1** and **2** were added to the set of 211 SNPs (5% FDR results and CAD fine-mapped regions), resulting in 153 independent SNPs (104 identified through the FDR analysis) at $r^2 < 0.2$, which were used for heritability calculations (**Supplementary Table 9**).

Heritability. Heritability estimates were calculated locus by locus using the multifactorial liability threshold model based on OR estimates that assume that the lead SNP at a locus accurately tags the disease-causing variant, as described in ref. 12. The calculations are based on a disease prevalence estimate of 5% and an estimate of 40% for the total heritability of coronary disease.

Expression analyses. We interrogated the 16 new (or proxy; $r^2 > 0.8$) CAD risk SNPs for *cis* eQTL expression in multiple tissues: the ASAP study[31] used tissue biopsies taken from patients undergoing carotid endartectomy (plaque $n = 117$) or valve surgery (liver $n = 152$, aorta media $n = 117$, aorta adventitia $n = 103$ and mammary artery $n = 88$). Expression data were generated using the Affymetrix HG-U133 plus 2.0 array (plaque) or the Affymetrix ST1.0 Exon array (liver, aorta and mammary artery); in the MuTHER study[32], RNA levels were measured in LCLs ($n = 826$), skin ($n = 705$) and fat biopsies ($n = 825$) from 850 female twins (one-third monozygotic and two-thirds dizygotic) from the TwinsUK resource using the Illumina HumanHT-12v3 array. We assessed genotype with gene expression associations, using an additive linear model (within a 1-Mb window); in Cardiogenics[5], monocytes and macrophages were collected from healthy subjects and individuals with CAD, and RNA was profiled with the Illumina Human Ref-8 array. eQTL analysis was undertaken in 459 healthy individuals from Cambridge, UK, using an additive linear model (1-Mb window); in the Massachusetts General Hospital study[33] of liver, omentum and subcutaneous adipose tissue among subjects undergoing Roux-en-Y gastric bypass surgery, eQTL analysis was performed with a linear regression model using a 1-Mb window.

In loci with significant *cis*-eQTL signal(s) ($P < 1 \times 10^{-4}$), we also identified the most strongly associated *cis*-eQTL SNP (eSNP) for the corresponding transcript and then performed conditional analyses, including in the regression model, with either the lead eSNP or the lead CAD-associated SNP. On the basis of the conditional analysis, we determined whether the same variant underlies both gene expression regulation and disease.

Finally, we interrogated the lead SNPs in the 16 new CAD susceptibility loci for allelic expression imbalance effects in LCLs, fibroblasts and monocytes ($n = 188$; Cardiogenics), as described in ref. 34.

Network analysis. Genes for network analysis were selected using 310 SNPs (88 SNPs in known and new CAD risk loci and 222 SNPs at FDR <10% and LD pruned to $r^2 \leq 0.7$). We first selected genes with an eQTL ($P \leq 1 \times 10^{-6}$) and then on the basis of physical proximity (included overlapping genes on opposite strands or at equal distance from the SNP; genes were considered within a 40-kb window centered on the SNP). Spliced ESTs and putative transcripts were not included. Network analysis was performed using the Ingenuity Pathway Analysis software tool (IPA; Ingenuity Systems). We considered molecules and/or relationships available in The IPA Knowledge Base for human, mouse or rat and set the confidence filter to experimentally observed or high (predicted). Networks were generated with a maximum size of 70 genes, allowing up to 10 networks. Molecules in the query set with recorded interactions were 'eligible' for network construction using the IPA algorithm. Networks were ranked according to their degree of relevance to the eligible molecules in the query data set. The score takes into account the number of eligible molecules in the network and its size, as well as the total number of eligible molecules

doi:10.1038/ng.2480

analyzed and the total number of molecules in the Ingenuity Knowledge Base that could potentially be included in the networks. The network score is based on the hypergeometric distribution and is calculated by right-tailed Fisher's exact test. The significance *P* value associated with enrichment of functional processes was calculated using the right-tailed Fisher's exact test by considering the number of query molecules that participate in that function and the total number of molecules that are known to be associated with that function in the Ingenuity Knowledge Base.

31. Folkersen, L. *et al.* Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ. Cardiovasc. Genet.* **3**, 365–373 (2010).
32. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins: the MuTHER Study. *Nat. Genet.* **44**, 1084–1089 (2012).
33. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
34. Ge, B. *et al.* Global patterns of *cis* variation in human cells revealed by high-density allelic expression analysis. *Nat. Genet.* **41**, 1216–1222 (2009).

Appendix 8: 153 SNPs included in the coronary artery disease gene score

SNP	CHR	Position	Nearest gene	Risk allele	Allele frequency	Weight
rs10965228	9	22072380	*9p21*	A	0.92	0.09
rs11238956	10	44069860	*CXCL12*	C	0.3	0.04
rs11617955	13	109616103	*COL4A1/COL4A2*	T	0.87	0.08
rs11619057	13	109806392	*COL4A2*	T	0.13	0.05
rs12127701	1	109639787	*MYBPHL*	A	0.94	0.07
rs12526453	6	13035530	*PHACTR1/RPEL*	C	0.68	0.08
rs12873154	13	109718853	*COL4A1/COL4A2*	G	0.12	0.07
rs16905599	9	22059144	*9p21*	G	0.89	0.08
rs17062853	6	134142738	*BC041459*	T	0.7	0.05
rs17155842	10	44072639	*CXCL12*	C	0.97	0.1
rs17465637	1	220890152	*MIA3*	C	0.72	0.05
rs17630235	12	111076069	*SH2B3/TRAFD1*	A	0.39	0.06
rs2244608	12	119901371	*HNF1A*	G	0.34	0.04
rs2288911	19	50141124	*APOC4/APOC2*	T	0.51	0.03
rs2521501	15	89238392	*FES*	T	0.31	0.06
rs2891403	12	111621955	*SH2B3/RPH3A*	A	0.26	0.05
rs3809274	12	110528716	*ATXN2*	G	0.78	0.06
rs4268379	1	109579760	*SARS*	T	0.49	0.04
rs6490029	12	110182840	*CUX2*	G	0.73	0.05
rs7139492	13	109713796	*COL4A1/COL4A2*	C	0.75	0.05
rs7181240	15	76932181	*MRG15*	G	0.15	0.05
rs7515901	1	109641419	*MYBPHL*	C	0.83	0.06
rs883947	6	12895700	*PHACTR1/RPEL*	C	0.67	0.04
rs892115	19	11124650	*SPC24*	G	0.32	0.04
rs9472428	6	12830159	*PHACTR1/RPEL*	A	0.41	0.04
rs9515201	13	109838799	*COL4A1/COL4A2*	A	0.3	0.05
rs11072794	15	76793637	*ADAMTS7/DQ582071*	T	0.28	0.07
rs11191447	10	104642313	*CYP17A1/CNNM2/NT5C2*	C	0.9	0.09
rs11203042	10	90979089	*LIPA*	T	0.45	0.04
rs11206510	1	55268627	*PCSK9*	T	0.82	0.06
rs1122608	19	11024601	*LDLR/SMARCA4*	G	0.76	0.09
rs11556924	7	129450732	*ZC3HC1*	C	0.64	0.08
rs12190287	6	134256218	*TCF21*	C	0.61	0.07
rs12205331	6	35006433	*ANKS1A*	C	0.79	0.04
rs12936587	17	17484447	*RAI1/PEMT/RASD1*	G	0.58	0.05
rs13211739	6	13070981	*PHACTR1*	G	0.25	0.06
rs1333049	9	22115503	*9p21*	C	0.47	0.21
rs15563	17	44360192	*UBE2Z*	G	0.53	0.04
rs17114036	1	56735409	*PPAP2B*	A	0.91	0.11
rs2047009	10	43859919	*CXCL12/AX747950*	G	0.5	0.05
rs2048327	6	160783522	*SLC22A3/LPAL2/LPA*	C	0.37	0.06
rs2075650	19	50087459	*APOE/APOC1/TOMM40*	G	0.14	0.11
rs2246833	10	90995834	*LIPA*	T	0.37	0.05

SNP	CHR	Position	Nearest gene	Risk allele	Allele frequency	Weight
rs2281727	17	2064695	*SMG6*	G	0.36	0.05
rs2306374	3	139602642	*MRAS*	C	0.16	0.07
rs2351524	2	203589237	*WDR12/ALS2CR16*	T	0.12	0.12
rs2505083	10	30375128	*KIAA1462*	C	0.42	0.06
rs2895811	14	99203695	*HHIPL1*	C	0.43	0.06
rs3184504	12	110368991	*SH2B3*	T	0.42	0.07
rs3217992	9	21993223	*CDKN2BAS/MTAP/CDKN2B*	T	0.38	0.15
rs445925	19	50107480	*APOE/APOC1/TOMM40*	G	0.9	0.12
rs4773144	13	109758713	*COL4A1/COL4A2*	G	0.43	0.07
rs495828	9	135144688	*ABO*	T	0.2	0.07
rs501120	10	44073873	*CXCL12*	T	0.84	0.07
rs602633	1	109623034	*PSRC1/SORT1*	G	0.78	0.12
rs7173743	15	76928839	*ADAMTS7/MRG15*	T	0.57	0.07
rs9326246	11	116116943	*ZNF259/APO5A/APOA1*	C	0.1	0.09
rs9515203	13	109847624	*COL4A1/COL4A2*	T	0.73	0.08
rs974819	11	103165777	*PDGFD*	T	0.29	0.06
rs9982601	21	34520998	*KCNE2/C21orf82*	T	0.14	0.12
rs10051876	5	87300520	*TMEM161B*	C	0.82	0.05
rs10237377	7	139403605	*PARP12/MST109*	G	0.66	0.05
rs1034565	22	18364211	*ARVCF*	T	0.31	0.04
rs10495907	2	43852230	*DYNC2LI1*	A	0.16	0.05
rs10507753	13	68180277	*BC042673*	T	0.06	0.07
rs10797416	1	2172202	*SKI*	T	0.62	0.04
rs10962774	9	16958831	*BNC2*	G	0.52	0.03
rs11204666	1	148809752	*MCL1*	C	0.82	0.05
rs1167800	7	75014132	*HIP1*	A	0.58	0.04
rs11710224	3	46561282	*LRRC2*	C	0.72	0.04
rs11718455	3	44031902	*DQ592230*	C	0.75	0.04
rs12125501	1	167540432	*NME7*	G	0.65	0.04
rs12663498	6	151045533	*PLEKHG1*	C	0.91	0.07
rs12801636	11	65147893	*PCNXL3*	G	0.77	0.05
rs1321309	6	36746614	*CDKN1A*	A	0.51	0.04
rs1393786	3	137336725	*PPP2R3A*	A	0.28	0.04
rs1490738	1	88910193	*PKN2*	A	0.39	0.03
rs16948048	17	44795465	*ZNF652*	G	0.35	0.03
rs17083481	4	54351705	*PDGFRA*	T	0.21	0.04
rs17087335	4	57533340	*C4orf14*	T	0.2	0.05
rs17318596	19	46628935	*ATP5SL*	A	0.38	0.04
rs17485781	8	27943481	*C8orf80*	C	0.94	0.1
rs17655141	4	44814329	*GNPDA2*	A	0.92	0.06
rs2071167	17	39643045	*UBTF*	C	0.76	0.04
rs217	7	27917546	*JAZF1*	T	0.64	0.04
rs2292096	1	199093392	*CAMSAP1L1*	G	0.12	0.06
rs2294461	6	6559500	*BC031936/AX746739/LY86*	C	0.1	0.06

SNP	CHR	Position	Nearest gene	Risk allele	Allele frequency	Weight
rs2395858	7	106751669	*COG5*	G	0.1	0.07
rs2571445	2	218391399	*TNS1*	A	0.4	0.04
rs2681472	12	88533090	*ATP2B1*	G	0.18	0.04
rs2736100	5	1339516	*TERT*	A	0.48	0.04
rs2832227	21	29454947	*C21orf7/DKFZp564A247*	G	0.16	0.05
rs2880765	15	83857466	*AKAP13*	A	0.48	0.04
rs4149033	12	21209077	*SLCO1B1*	G	0.77	0.04
rs4301033	3	151525308	*TSC22D2*	A	0.09	0.06
rs4410190	18	18274198	*CTAGE1*	T	0.47	0.03
rs4469055	4	148605171	*EDNRA*	G	0.84	0.05
rs4566357	2	227630259	*COL4A4*	A	0.38	0.04
rs4591971	7	130996735	*PODXL*	A	0.26	0.04
rs4690974	4	156613091	*MAP9*	C	0.48	0.04
rs4762911	12	20052013	*PDE3A*	A	0.74	0.04
rs4793721	17	47166312	*CA10*	C	0.45	0.03
rs6088638	20	32934175	*ACSS2*	C	0.16	0.05
rs683800	11	125688966	*DCPS*	T	0.56	0.03
rs6926458	6	160939856	*LPA*	A	0.75	0.05
rs7074064	10	88673102	*BMPR1A*	C	0.27	0.04
rs7116641	11	43653493	*HSD17B12*	G	0.31	0.04
rs7356185	4	120386559	*USP53*	T	0.23	0.04
rs7496815	15	89862501	*BC036442*	A	0.44	0.04
rs7561273	2	24101018	*LOC388931*	A	0.46	0.03
rs7642590	3	48074754	*MAP4/FKSG52*	A	0.68	0.04
rs816889	2	151033541	*RND3*	G	0.82	0.05
rs93139	11	9716184	*SWAP70*	C	0.59	0.04
rs972158	7	26301532	*SNX10*	T	0.57	0.04
rs10947789	6	39282900	*KCNK5*	T	0.76	0.06
rs1429141	4	148507517	*EDNRA*	T	0.81	0.07
rs1561198	2	85663500	*GGCX/VAMP10/VAMP8*	T	0.45	0.05
rs16986953	2	19805954	*AK097927*	A	0.09	0.1
rs17514846	15	89217554	*FES/FURIN*	A	0.44	0.06
rs2023938	7	19003300	*HDAC9*	C	0.11	0.07
rs2252641	2	145517931	*ZEB2-AC074093.1*	C	0.45	0.05
rs264	8	19857460	*LPL*	G	0.85	0.07
rs273909	5	131695252	*SLC22A4/SLC22A5*	G	0.13	0.08
rs2954029	8	126560154	*TRIB1*	A	0.54	0.05
rs4252120	6	161063598	*PLG*	T	0.72	0.06
rs4845625	1	152688691	*IL6R*	T	0.44	0.05
rs515135	2	21139562	*APOB*	C	0.83	0.07
rs6544713	2	43927385	*ABCG5/ABCG8*	T	0.29	0.06
rs7692387	4	156854759	*GUCY1A3*	G	0.8	0.07
rs9319428	13	27871621	*FLT1*	A	0.32	0.06
rs11057841	12	123882696	*SCARB1*	T	0.15	0.07

SNP	CHR	Position	Nearest gene	Risk allele	Allele frequency	Weight
rs11806316	1	115555005	*NGF*	G	0.63	0.04
rs11916151	3	88363366	*C3orf38*	T	0.92	0.11
rs1247351	6	161283909	*PLG/MAP3K4*	C	0.3	0.05
rs12765878	10	105659612	*OBFC1*	C	0.5	0.04
rs2070783	17	59760703	*PECAM1*	G	0.54	0.04
rs2146238	14	99242482	*CYP46A1*	G	0.84	0.06
rs246600	5	142497090	*ARHGAP26*	T	0.48	0.05
rs2820315	1	200138887	*LMOD1*	T	0.31	0.05
rs3748242	10	81904767	*ANXA11*	T	0.24	0.05
rs3778448	6	39271179	*KCNK5*	G	0.33	0.05
rs4299203	17	17818884	*TOM1L2/LRRC48/ATPAF2*	G	0.37	0.05
rs4613862	6	82668990	*BC038576*	A	0.53	0.04
rs590121	11	74951798	*SERPINH1*	T	0.3	0.05
rs606452	11	74953826	*SERPINH1*	C	0.86	0.06
rs6494488	15	62811257	*RBPMS2*	A	0.82	0.06
rs6700559	1	198912696	*DDX59*	C	0.54	0.04
rs6841581	4	148620640	*EDNRA*	A	0.16	0.07
rs6984210	8	22089560	*BMP1*	G	0.07	0.1
rs8111989	19	50501048	*MARK4/CKM*	C	0.31	0.05
rs867186	20	33228215	*PROCR*	A	0.89	0.07
rs9316753	13	54365930	*BC044614*	C	0.63	0.04
rs9608859	22	28997277	*OSM*	C	0.58	0.05

Appendix 9. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): meta-analysis of genome-wide association studies

Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies

Matthew Traylor, Martin Farrall, Elizabeth G Holliday, Cathie Sudlow, Jemma C Hopewell, Yu-Ching Cheng, Myriam Fornage, M Arfan Ikram, Rainer Malik, Steve Bevan, Unnur Thorsteinsdottir, Mike A Nalls, W T Longstreth, Kerri L Wiggins, Sunaina Yadav, Eugenio A Parati, Anita L DeStefano, Bradford B Worrall, Steven J Kittner, Muhammad Saleem Khan, Alex P Reiner, Anna Helgadottir, Sefanja Achterberg, Israel Fernandez-Cadenas, Sherine Abboud, Reinhold Schmidt, Matthew Walters, Wei-Min Chen, E Bernd Ringelstein, Martin O'Donnell, Weang Kee Ho, Joanna Pera, Robin Lemmens, Bo Norrving, Peter Higgins, Marianne Benn, Michele Sale, Gregor Kuhlenbäumer, Alexander S F Doney, Astrid M Vicente, Hossein Delavaran, Ale Algra, Gail Davies, Sofia A Oliveira, Colin N A Palmer, Ian Deary, Helena Schmidt, Massimo Pandolfo, Joan Montaner, Cara Carty, Paul I W de Bakker, Konstantinos Kostulas, Jose M Ferro, Natalie R van Zuydam, Einar Valdimarsson, Børge G Nordestgaard, Arne Lindgren, Vincent Thijs, Agnieszka Slowik, Danish Saleheen, Guillaume Paré, Klaus Berger, Gudmar Thorleifsson, The Australian Stroke Genetics Collaborative, Wellcome Trust Case Control Consortium 2 (WTCCC2), Albert Hofman, Thomas H Mosley, Braxton D Mitchell, Karen Furie, Robert Clarke, Christopher Levi, Sudha Seshadri, Andreas Gschwendtner, Giorgio B Boncoraglio, Pankaj Sharma, Joshua C Bis, Solveig Gretarsdottir, Bruce M Psaty, Peter M Rothwell, Jonathan Rosand, James F Meschia, Kari Stefansson, Martin Dichgans, Hugh S Markus, on behalf of the International Stroke Genetics Consortium

Summary

Background Various genome-wide association studies (GWAS) have been done in ischaemic stroke, identifying a few loci associated with the disease, but sample sizes have been 3500 cases or less. We established the METASTROKE collaboration with the aim of validating associations from previous GWAS and identifying novel genetic associations through meta-analysis of GWAS datasets for ischaemic stroke and its subtypes.

Methods We meta-analysed data from 15 ischaemic stroke cohorts with a total of 12 389 individuals with ischaemic stroke and 62 004 controls, all of European ancestry. For the associations reaching genome-wide significance in METASTROKE, we did a further analysis, conditioning on the lead single nucleotide polymorphism in every associated region. Replication of novel suggestive signals was done in 13 347 cases and 29 083 controls.

Findings We verified previous associations for cardioembolic stroke near *PITX2* ($p=2\cdot8\times10^{-16}$) and *ZFHX3* ($p=2\cdot28\times10^{-8}$), and for large-vessel stroke at a 9p21 locus ($p=3\cdot32\times10^{-5}$) and *HDAC9* ($p=2\cdot03\times10^{-12}$). Additionally, we verified that all associations were subtype specific. Conditional analysis in the three regions for which the associations reached genome-wide significance (*PITX2*, *ZFHX3*, and *HDAC9*) indicated that all the signal in each region could be attributed to one risk haplotype. We also identified 12 potentially novel loci at $p<5\times10^{-6}$. However, we were unable to replicate any of these novel associations in the replication cohort.

Interpretation Our results show that, although genetic variants can be detected in patients with ischaemic stroke when compared with controls, all associations we were able to confirm are specific to a stroke subtype. This finding has two implications. First, to maximise success of genetic studies in ischaemic stroke, detailed stroke subtyping is required. Second, different genetic pathophysiological mechanisms seem to be associated with different stroke subtypes.

Funding Wellcome Trust, UK Medical Research Council (MRC), Australian National and Medical Health Research Council, National Institutes of Health (NIH) including National Heart, Lung and Blood Institute (NHLBI), the National Institute on Aging (NIA), the National Human Genome Research Institute (NHGRI), and the National Institute of Neurological Disorders and Stroke (NINDS).

Lancet Neurol 2012; 11: 951–62

Published Online
October 5, 2012
http://dx.doi.org/10.1016/S1474-4422(12)70234-X

See **Comment** page 931

Authors' affiliations listed at end of paper

Correspondence to:
Dr Hugh S Markus, Stroke and Dementia Research Centre, St George's University of London, Cranmer Terrace, London, SW17 0RE, UK
hmarkus@sgul.ac.uk

Introduction

Stroke is one of the three most common causes of death, is a major cause of adult chronic disability,[1] and represents an important cause of age-related cognitive decline and dementia. Conventional risk factors explain only a small proportion of all stroke risk.[2] Evidence from studies of twins and family history suggests that genetic predisposition is important.[3] In common with many other complex diseases, in which environmental risk factors are thought to interact with multiple genes, the identification of the underlying molecular mechanisms contributing to stroke risk has been a challenge. Candidate gene studies have produced few replicable associations.[4] More recently, the genome-wide association study (GWAS) approach has transformed the genetics of other complex diseases and is just beginning to affect the study of stroke.[5,6]

About 80% of stroke is ischaemic, whereas 20% is due to primary haemorrhage.[6] Ischaemic stroke itself includes several subtypes with differing pathophysiological mechanisms, the most common of which are large-vessel disease stroke, small-vessel disease stroke, and cardioembolic stroke.[7] Various genetic

variants that predispose to risk factors for stroke have also been shown in GWAS to predispose to ischaemic stroke.[8–10] Two loci associated with atrial fibrillation (*PITX2* and *ZFHX3*) were associated with cardioembolic stroke, whereas a locus on chromosome 9p21 originally associated with coronary artery disease was shown to be a risk factor for large-vessel stroke.[8–10] The few novel stroke-associated loci reported to date have been mainly associated with stroke subtypes, rather than with the phenotype of ischaemic stroke. In Japanese populations, a variant in the protein kinase C family (*PRKCH*) was associated with small-vessel stroke.[11] A meta-analysis of prospective population-based cohort studies reported an association with the 12p13 region, thought to be with the *NINJ2* gene, although this result was not replicated in a larger case-control sample.[12,13] Recently, the Wellcome Trust Case Control Consortium 2 (WTCCC2) GWAS in ischaemic stroke reported a novel association on chromosome 7p21 within the *HDAC9* gene, although it was associated only with large-vessel ischaemic stroke.[14]

GWAS in ischaemic stroke to date have used small discovery populations, with the largest including 3548 individuals.[14] In other complex diseases, many additional associations have been detected as the discovery sample size has increased.[15–17] This increase has usually been achieved by meta-analysis of independent datasets. Therefore, we established the METASTROKE collaboration to combine the available GWAS datasets of ischaemic stroke. Here, we describe the first paper from METASTROKE with a description of the constituent cohorts. Using this dataset, we attempted both to replicate previous GWAS associations with ischaemic stroke and to identify novel associations. Additionally, we determined whether stroke loci were specific to individual stroke subtypes.

Methods

Study design and participating studies

The discovery sample consisted of 15 cohorts of patients with ischaemic stroke who were of European ancestry from Europe, North America, and Australia, together with controls of matched ancestry. All studies used a case-control methodology. Most participating studies were cross-sectional, whereas four were in large, prospective, population-based cohorts (table 1).

Additionally, 18 cohorts were analysed in the replication phase. These cohorts were included for replication only, most did not have GWAS data available; and those with GWAS data were not available at the time of the discovery analysis. 17 of the included cohorts contained individuals of solely European ancestry, and one contained individuals of Pakistani ancestry (table 1). Most cohorts (16) were cross-sectional, whereas two were population-based.

See Online for appendix

The appendix includes detailed descriptions of the design and clinical characteristics of the participating studies.

Stroke was defined as a typical clinical syndrome with radiological confirmation. Stroke subtyping was done with the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification system.[18] Where subtyping was done, brain CT or MRI was undertaken for more than 95% of cases in all the discovery cohorts.

Participating studies were approved by relevant institutional review boards, and all participants gave written or oral consent for study participation, including genetic research, as approved by the local institutional body.

Data imputation and statistical analysis

The 15 discovery cohorts used commercially available GWAS panels of single nucleotide polymorphisms (SNPs) from either Affymetrix (Santa Clara, CA, USA) or Illumina (San Diego, CA, USA). 14 of the 15 centres undertook genotype imputation with HapMap II,[19] HapMap III,[20] or 1000 Genomes[21] as reference haplotype training sets. Every centre did genotypic quality control steps before imputation, including removal of ancestry outliers defined by principal component analysis and poorly typed individuals.

We used logistic regression for all cohorts with a cross-sectional study design to model the multiplicative SNP effects on risk for the dichotomous outcome of stroke against ancestry-matched controls, whereas we used Cox proportional-hazards models for the prospective studies to assess time to first stroke, fitting an additive model relating genotype dose to the stroke outcome. Where genotypes were imputed, SNPs were modelled as allele dosages. Of the discovery cohorts, four (of 15) centres used ancestry-informative principal components as covariates to correct for population stratification. All cohorts providing genome-wide data removed population outliers before imputation. After verifying strand alignment, filtering SNPs with minor allele frequency lower than 0·01, and removing poorly imputed SNPs across centres, we did a meta-analysis of the results of the association analyses from every centre using a fixed-effects inverse-variance weighted model using METAL.[22]

We sought further evidence for association with novel suggestively associated SNPs in new samples from 18 different cohorts. Of the 18 centres, six submitted in-silico genotype data and 12 undertook direct genotyping with the Sequenom (Sequenom, San Diego, CA, USA) or Taqman (Applied Biosystems, Foster City, CA, USA) platforms. All of the five replication cohorts contributing genome-wide data used principal components as covariates in their analyses. We did a meta-analysis of the results for the replication cohorts using a fixed-effects, inverse variance weighted method first for all datasets, and then for replication datasets of solely European ancestry. We determined whether SNPs were significantly associated in the replication population, and additionally, we combined results from the discovery and replication analyses using a fixed-effects, inverse-variance weighted approach.

	Number of cases	Number of CS cases	Number of LVD cases	Number of SVD cases	Number of controls	Ancestry	Study design	Genotyping
Discovery cohorts								
ARIC	385	93	31	63	8803	European	Population-based	Affymetrix 6.0
ASGC	1162	240	421	310	1244	European	Cross-sectional	Illumina 610
BRAINS	361	29	120	97	444	European	Cross-sectional	Illumina 660
CHS	454	147	..	73	2817	European	Population-based	Illumina 370
deCODE	2391	399	255	240	26 970	European	Cross-sectional	Illumina 317/370
FHS	171	48	..	..	4164	European	Population-based	Affymetrix 550
GEOS	448	90	37	54	498	European	Cross-sectional	Illumina HumanOmni1
HPS	578	..	..	..	468	European	Cross-sectional	Illumina 610
HVH	566	88	61	173	1290	European	Cross-sectional	Illumina 370
ISGS/SWISS	1070	247	229	201	2329	European	Cross-sectional	Illumina 550/610/660
MGH-GASROS	516	169	95	38	1202	European	Cross-sectional	Affymetrix 6.0
Milano	372	25	74	65	407	European	Cross-sectional	Illumina 610/660
Rotterdam	367	..	..	..	5396	European	Population-based	Illumina 550
WTCCC2-Munich	1174	330	346	106	797	European	Cross-sectional	Illumina 660
WTCCC2-UK	2374	460	498	474	5175	European	Cross-sectional	Illumina 660
Total (discovery)	12 389	2365	2167	1894	62 004	..	..	..
Replication cohorts								
Barcelona	439	179	110	150	404	European	Cross-sectional	Sequenom
BSS	225	11	93	90	312	European	Cross-sectional	Sequenom
Copenhagen	730	..	..	..	1545	European	Cross-sectional	TaqMan
ESS	276	40	20	69	940	European	Cross-sectional	TaqMan/Illumina 610
Glasgow	675	125	91	150	940	European	Cross-sectional	Sequenom/Illumina 610
Go-Darts*	737	130	259	..	8424	European	Cross-sectional	Affymetrix 6.0/Illumina Cardio-metabochip
Graz	657	116	108	207	848	European	Cross-sectional	Sequenom/Illumina 610
Interstroke*	872	143	198	238	926	European	Cross-sectional	Illumina Cardio-metabochip
Krakow	1235	377	152	171	584	European	Cross-sectional	Sequenom
Leuven	458	195	83	63	391	European	Cross-sectional	Sequenom
Lund	424	140	21	94	466	European	Cross-sectional	Sequenom
Munster	1232	478	528	224	1053	European	Cross-sectional	Sequenom
Portugal	539	..	..	..	507	European	Cross-sectional	Sequenom
RACE (Pakistan)*	1322	225	195	189	1143	Pakistani	Cross-sectional	Illumina 660
SMART	623	30	368	195	6712	European	Population-based	Sequenom
Sweden	876	157	177	75	742	European	Cross-sectional	Sequenom
VISP*	1725	..	..	..	1047	European	Cross-sectional	Illumina HumanOmni1
WHI*	302	42	31	78	2099	European	Population-based	Illumina Omni-Quad
Total (replication)	13 347	2388	2434	1993	29 083	..	..	..

CS=cardioembolic stroke. LVD=large-vessel disease. SVD=small-vessel disease. ARIC=The Atherosclerosis Risk in Communities study. ASGC=Australian Stroke Genetics Collabarative. BRAINS=Bio-Repository of DNA in stroke. CHS=Cardiovascular Health Study. FHS=Framingham Heart Study. GEOS=Genetics of Early-Onset Stroke. HPS=Heart Protection Study. HVH=The Heart and Vascular Health Study. ISGS/SWISS=The Ischemic Stroke Genetics Study/Sibling with Ischaemic Stroke Study. MGH-GASROS=The MGH Genes Affecting Stroke Risk and Outcome Study. WTCCC2-Munich=The Wellcome Trust Case-Control Consortium II Munich. WTCCC2-UK=The Wellcome Trust Case-Control Consortium II UK. BSS=Belgium Stroke Study. ESS=Edinburgh Stroke Study. Go-Darts=Genetics of Diabetes Audit and Research in Tayside Study. RACE=Risk Assessment of Cerebrovascular Events Study, Pakistan. SMART=Second Manifestations of ARTerial disease. VISP=The Vitamin Intervention for Stroke Prevention Trial. WHI=The Women's Health Initiative. *Contributed genome-wide data.

Table 1: **Description of cohorts used in analysis by study population**

We set the study-wide genome-wide significance level at $p<5\times10^{-8}$ to control the experiment-wide error rate to <5%. Following the example of previous GWAS studies,[15] we set the level for suggestive significance at $p<5\times10^{-6}$.

First, we attempted to determine the evidence for association for the six loci reported previously from GWAS to be associated with ischaemic stroke (*HDAC9, PITX2, ZFHX3, NINJ2, PRKCH,* and 9p21).[8–12,14] After determining the evidence for association with the previously reported SNPs, we investigated whether any proxy SNPs were more significantly associated in the METASTROKE dataset. Because some loci had been identified in discovery populations included in METASTROKE, we initially did analyses for the whole dataset, and then we restricted

analysis to the lead SNP for every locus in the METASTROKE cohorts that had not been included in the discovery phase of the initial publication. We set the significance level for independent replication at $p<0·01$, corresponding to Bonferroni corrected type 1 error <5% for the five SNPs (excluding *PRKCH*) tested.

As the SNP in *PRKCH* (rs2230500) underlying the previous association in Japanese cohorts[5] is monomorphic in populations of European ancestry, we sought to identify any associations within this gene region, including the 50 kbp window upstream and downstream, in our large population of European ancestry. Using the modified Nyholt correction approach of Li and Ji on the 353 SNPs from the region, we estimated the effective number of SNPs tested to be 103·3.[23] We therefore set the significance level at $p<0·00048$, corresponding to Bonferroni corrected type I error <5% for the effective SNPs tested.

We also did an analysis to determine whether the six previously reported variants were associated with stroke risk in prospective population-based studies. We did this analysis only for the known SNPs that had been analysed in a minimum of 100 cases in the prospective cohorts with incident stroke events for the relevant subtype.

For those associations we could confirm, we then did a conditional analysis within the associated region to identify any signal in the region that was independent of the lead SNP in every case. For every association, we selected regions used in the conditional analysis on the basis of adjacent recombination hotspots, meaning we analysed different numbers of SNPs for every locus (appendix). We used logistic regression in every centre, using imputed genotype dosages to model the effect of the lead SNP on risk as a covariate. We then did a meta-analysis of the results using a fixed-effects, inverse-variance weighted model. We used our suggestive significance threshold ($p<5\times10^{-6}$) to identify SNPs that were statistically independent of the lead SNP for every locus.

We then did a meta-analysis of the genome-wide study-specific analysed datasets to identify novel associations with ischaemic stroke and its subtypes. We did the primary association analyses for all ischaemic stroke and for the three major subtypes: cardioembolic stroke, large-vessel disease, and small-vessel disease. We did additional secondary analyses for young cases (younger than 70 years at first stroke) and for the phenotype of ischaemic stroke in each sex separately. We reused the same controls per centre for all analyses. Excluding the previously published associations, we considered all SNPs reaching suggestive significance ($p<5\times10^{-6}$) for replication. We examined SNPs for heterogeneity across datasets and attempted replication in independent datasets for the loci that were deemed plausible candidates for association with ischaemic stroke.

For a minor allele frequency of 0·25, we had 80% power to detect variants with a per-allele odds ratio (OR) greater than 1·11 for the all ischaemic stroke analysis, 1·23 for cardioembolic stroke, 1·24 for large-vessel disease, and 1·26 for small-vessel disease at $p<5\times10^{-8}$ in the discovery phase.

Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The discovery meta-analysis of ischaemic stroke phenotypes involved a total of 12 389 cases and 62 004 controls from 15 populations (table 1; figure 1).

The discovery meta-analysis confirmed associations at genome-wide significance levels for *HDAC9* with large-vessel disease, and for both *PITX2* and *ZFHX3* with cardioembolic stroke (table 2). For *PITX2*, *ZFHX3*, and *HDAC9* a proxy SNP was more significant in the METASTROKE dataset than the SNP from the original publication (original SNP shown in appendix). The 9p21 locus was associated with large-vessel disease with a similar OR (1·15, 95% CI 1·08–1·23, in METASTROKE) to that reported previously (1·21, 1·07–1·37),[10] although it did not reach genome-wide significance ($p=3·32\times10^{-5}$). All four associations were subtype specific, being present only for a single stroke subtype (table 2). To determine the extent to which these results replicated the findings from the originally published associations, we repeated the meta-analysis, this time excluding the populations that contributed to the discovery phase of the original publication. For the *PITX2*, *ZFHX3*, *HDAC9*, and 9p21 loci, the associations were replicated in the independent METASTROKE samples (table 2). The population attributable risks in the METASTROKE discovery cohort were estimated as 5·8% for *PITX2* and 7·0% for *ZFHX3* in cardioembolic stroke, and 4·5% for *HDAC9* and 7·2% for 9p21 in large-vessel disease.

The *NINJ2* locus showed nominal evidence of association with all ischaemic stroke when all populations were included (table 2). However, no evidence was noted for association with the *NINJ2* locus when the original discovery populations were excluded (table 2).

To estimate the effect of these associations in prospective population-based studies, we had a sufficient number of stroke cases for the analysis in only the cardioembolic subtype (n=376). We noted ORs similar to those identified in the overall case-control study for both *PITX2* (1·26, 95% CI 1·05–1·52, in prospective studies and 1·36, 1·27–1·47, in case-control analysis) and *ZFHX3* (1·23, 0·98–1·55, in prospective studies and 1·25, 1·15–1·35, in case-control analysis), although this similarity was significant only for *PITX2* (appendix).

We found no significant associations between the *PRKCH* gene region and all ischaemic strokes or with the three main subtype analyses. Table 2 provides details

	Chr	BP	SNP	RA	RAF	Full METASTROKE discovery sample		Excluding cohorts used in previous discovery of relevant association*	
						OR (95% CI)	p value†	OR (95% CI)	p value†
HDAC9	7	19 015 913	rs2107595	A	0·16				
IS	..	..	..	..	..	1·12 (1·07–1·17)	$4·34×10^{-6}$	1·11 (1·05–1·17)	$7·8×10^{-5}$
LVD	..	..	..	..	..	1·39 (1·27–1·53)	$2·03×10^{-14}$	1·39 (1·24–1·56)	$3·15×10^{-8}$
SVD	..	..	..	..	..	1·03 (0·93–1·14)	0·57	1·11 (0·96–1·29)	0·92
CE	..	..	..	..	..	1·07 (0·98–1·17)	0·15	1·07 (0·96–1·19)	0·25
PITX2	4	111 937 516	rs6843082	G	0·21				
IS	..	..	..	..	..	1·11 (1·06–1·15)	$1·95×10^{-7}$	1·09 (1·04–1·14)	$1·12×10^{-4}$
LVD	..	..	..	..	..	1·06 (0·97–1·15)	0·17	1·03 (0·93–1·13)	0·61
SVD	..	..	..	..	..	1·04 (0·96–1·14)	0·31	1·01 (0·90–1·13)	0·91
CE	..	..	..	..	..	1·36 (1·27–1·47)	$2·8×10^{-16}$	1·32 (1·23–1·44)	$3·64×10^{-12}$
ZFHX3	16	71 626 169	rs879324	A	0·19				
IS	..	..	..	..	..	1·05 (1·00–1·09)	0·037	1·06 (1·01–1·11)	0·021
LVD	..	..	..	..	..	1·06 (0·98–1·16)	0·15	1·06 (0·96–1·17)	0·32
SVD	..	..	..	..	..	0·99 (0·91–1·09)	0·94	1·01 (0·91–1·13)	0·81
CE	..	..	..	..	..	1·25 (1·15–1·35)	$2·28×10^{-8}$	1·25 (1·15–1·36)	$1·53×10^{-7}$
NINJ2	12	645 460	rs11833579	A	0·22				
IS	..	..	..	..	..	1·06 (1·02–1·10)	$6·1×10^{-4}$	1·00 (0·96–1·05)	0·81
LVD	..	..	..	..	..	0·99 (0·91–1·08)	0·87	0·99 (0·91–1·08)	0·79
SVD	..	..	..	..	..	0·98 (0·90–1·08)	0·79	0·99 (0·90–1·08)	0·79
CE	..	..	..	..	..	1·04 (0·97–1·13)	0·27	1·00 (0·92–1·09)	0·95
9p21	9	22 105 959	rs2383207	G	0·52				
IS	..	..	..	..	..	1·04 (0·76–1·41)	0·024	1·03 (0·99–1·07)	0·16
LVD	..	..	..	..	..	1·15 (1·08–1·23)	$3·32×10^{-6}$	1·15 (1·04–1·27)	$5·69×10^{-3}$
SVD	..	..	..	..	..	1·02 (0·96–1·10)	0·48	1·03 (0·93–1·14)	0·61
CE	..	..	..	..	..	0·96 (0·91–1·03)	0·24	1·02 (0·92–1·14)	0·61
PRKCH									
IS	14	61 077 900	rs2246700	A	0·84	1·07 (1·02–1·12)	0·0049	..	..
LVD	14	60 894 555	rs12587610	G	0·31	1·11 (1·03–1·21)	0·0046	..	..
SVD	14	61 114 037	rs2255146	G	0·82	1·22 (1·03–1·43)	0·0175	..	..
CE	14	60 988 886	rs3825655	C	0·95	1·31 (1·00–1·71)	0·0475	..	..

Chr=chromosome. BP=base position. SNP=single nucleotide polymorphism. RA=risk allele. RAF=risk allele frequency. OR=odds ratio. IS=all ischaemic strokes. LVD=large vessel disease. SVD=small vessel disease. CE=cardioembolic stroke. *Statistics shown are after removal of discovery populations showing an association between the gene and stroke from original publications—ie, deCODE excluded for *PITX2*, *ZFHX3*;[8,9] WTCCC2-UK and WTCCC-Munich excluded for *HDAC9*;[14] WTCCC2-UK and WTCCC2-Munich, ISGS/SWISS, GEOS, and MGH-GASROS excluded for *CDKN2a/CDKN2b* (9p21);[10] Rotterdam, ARIC, FHS, and CHS excluded for *NINJ2*.[11] †One-sided p value.

Table 2: **METASTROKE association signals for SNPs identified in previous genome-wide association studies by gene and disease subtype**

of the most strongly associated SNPs in every subtype for this locus.

For those loci for which we confirmed genome-wide significance (*PITX2*, *ZFHX3*, and *HDAC9*), we did conditional analyses. After conditioning on the lead SNP in the given region, no SNP showed significance at $p<0·01$ in *PITX2* or *ZFHX3*, and no SNP showed significance at $p<0·005$ in *HDAC9*. Furthermore, all other SNPs in the regions that were associated at $p<5×10^{-8}$ in the main analysis showed no significance ($p>0·05$) in any of the analyses after conditioning on the lead SNP. Figure 2 shows plots of $-\log_{10}$(p values) against genomic position in the selected regions for the unconditional and conditional analyses.

We selected a total of 12 novel SNPs for testing in the independent replication cohort: three associated with all

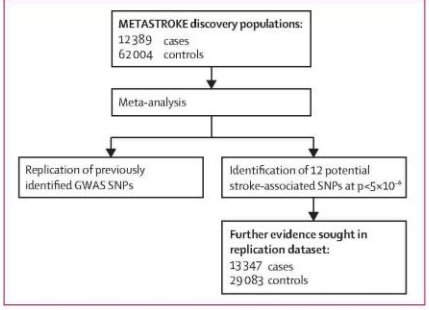

Figure 1: **Flow diagram of METASTROKE analyses**
GWAS=genome-wide association study. SNP=single nucleotide polymorphism.

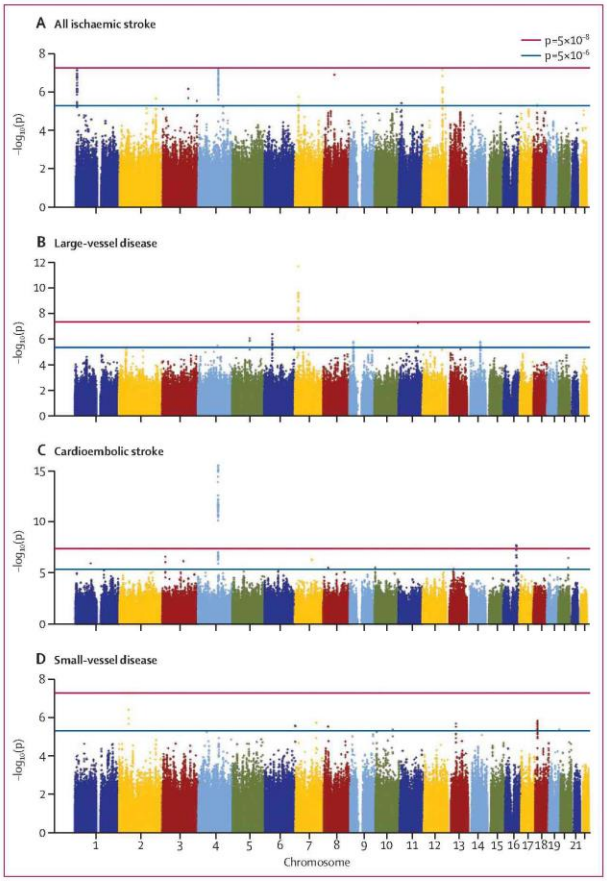

Figure 2: **Manhattan plots of $-\log_{10}(p)$ against genomic position for principal analyses**
(A) All ischaemic stroke. (B) Large-vessel disease. (C) Cardioembolic stroke. (D) Small-vessel disease. Genome-wide meta-analysis association results by genomic position for the four main analyses.

ischaemic stroke, five associated with specific stroke subtypes, and two each associated with young stroke and female stroke. Four of these SNPs showed associations close to genome-wide significance in the discovery cohort: rs225132 in the *ERRF11* gene and rs17696736 in the *NAA25* (C12orf30) gene with all ischaemic stroke ($p=6\cdot3\times10^{-8}$ and $5\cdot9\times10^{-8}$, respectively), rs7937106 in *ALKBH8* with large-vessel disease ($p=5\cdot9\times10^{-8}$), and rs13407662 on chromosome 2p16.2 ($p=5\cdot2\times10^{-8}$) in an intergenic region with small-vessel disease. The remaining SNPs were identified at the suggestive significance level of $p<5\times10^{-6}$. Table 3 shows details of these SNPs, including stroke subtypes with which they were associated, and significance levels. These 12 novel SNPs were taken forward for replication in an additional 13 347 cases and 29 083 controls. Figure 3 shows the plots of $-\log_{10}$(p values) by chromosomal location for the analysis of all stroke and the three main subtypes.

None of the novel SNPs reached genome-wide significance on combination of the discovery and replication data. This result was the same when replication analysis was restricted to individuals of European ancestry (table 3). There was significant heterogeneity ($p<0\cdot05$) for all of the SNPs in the combined analysis. We had sufficient sample size to obtain 80% power to confirm each of the 12 loci (appendix).

Discussion

METASTROKE is the first large meta-analysis of stroke GWAS data (panel). The METASTROKE collaboration brings together GWAS data from more than 12 000 cases of ischaemic stroke and 60 000 controls from 15 cohorts all of European ancestry. In this first analysis from the dataset, we confirmed four of five previously described associations with ischaemic stroke in populations of European ancestry, including replication in an independent non-overlapping sample of the dataset not included in the original GWAS. All these associations were with specific subtypes of ischaemic stroke, emphasising the genetic heterogeneity of the disease. Additionally, we identified several promising novel associations, some of which were close to genome-wide significance in the discovery cohorts, but these were not confirmed in our replication population.

Our results provide further robust data supporting an association between two gene regions (*PITX2* and *ZFHX3*) and cardioembolic stroke, and a further two (*HDAC9* and 9p21) with large-vessel stroke although the 9p21 locus did not reach genome-wide significance. In all cases, these associations were present in the dataset as a whole, and also when those samples used in the original discovery cohorts that identified associations with ischaemic stroke were excluded.

Both *PITX2* and *ZFHX3* were originally identified as risk factors for atrial fibrillation.[8,9] Atrial fibrillation is a major risk factor for stroke, particularly in the elderly, and therefore their association with ischaemic stroke is not unexpected. Our results confirm this association and clearly show that it is limited to the cardioembolic stroke subtype. Furthermore, we were able to show an association between *PITX2* and ischaemic stroke in prospective cohorts. A potential bias is that a variant that is in fact associated with mortality rate after acute stroke and not with stroke risk might seem to be related to risk; for cross-sectional studies in a disease such as stroke, which has substantial early mortality, death might occur before or soon after hospital admission before samples are taken. In a prospective study, such cases are included

	Chr	SNP	Candidate gene	RA	RAF	$p_{discovery}$; $OR_{discovery}$ (95% CI)	All replication samples		Replication in European descent individuals only	
							$p_{replication}$; $OR_{replication}$ (95% CI)	$p_{combined}$	$p_{replication}$; $OR_{replication}$ (95% CI)	$p_{combined}$
IS	1	rs225132	*ERRF11*	T	0·82	$6·27\times10^{-8}$; 1·12 (1·07–1·17)	0·16; 0·97 (0·92–1·01)	$1·65\times10^{-3}$	0·11; 0·96 (0·92–1·01)	$1·91\times10^{-3}$
IS	12	rs17696736	*NAA25 (C12orf30)*	G	0·42	$5·97\times10^{-8}$; 1·10 (1·06–1·14)	0·59; 1·01 (0·97–1·05)	$1·92\times10^{-5}$	0·60; 1·01 (0·97–1·05)	$1·69\times10^{-5}$
IS	3	rs16851055	*SPSB4*	G	0·81	$6·34\times10^{-7}$; 1·12 (1·07–1·17)	0·20; 1·03 (0·98–1·08)	$6·23\times10^{-4}$	0·25; 1·03 (0·98–1·08)	$7·76\times10^{-4}$
CS	3	rs6763538	*OXNAD1*	T	0·04	$2·89\times10^{-7}$; 1·47 (1·27–1·69)	0·69; 1·04 (0·87–1·24)	$2·68\times10^{-4}$	0·59; 1·05 (0·88–1·25)	$1·36\times10^{-4}$
LVD	11	rs7937106	*ALKBH8*	C	0·16	$5·85\times10^{-8}$; 1·68 (1·40–2·03)	0·66; 1·04 (0·87–1·25)	$3·93\times10^{-4}$	0·65; 1·05 (0·85–1·31)	$1·42\times10^{-4}$
LVD	6	rs556621	..	T	0·33	$4·63\times10^{-7}$; 1·20 (1·12–1·28)	0·46; 1·03 (0·96–1·10)	$5·33\times10^{-4}$	0·37; 1·03 (0·96–1·11)	$2·43\times10^{-4}$
SVD	18	rs7407640	*AFG3L2*	A	0·21	$2·20\times10^{-6}$; 1·23 (1·13–1·34)	0·99; 1·00 (0·91–1·10)	$4·54\times10^{-4}$	0·57; 0·97 (0·88–1·07)	$1·16\times10^{-3}$
SVD	2	rs13407662	..	T	0·04	$5·18\times10^{-8}$; 1·95 (1·53–2·48)	0·28; 1·16 (0·89–1·51)	$1·97\times10^{-4}$	0·36; 1·14 (0·86–1·53)	$1·88\times10^{-4}$
FS	3	rs7432308	..	T	0·15	$1·63\times10^{-6}$ 1·16 (1·09–1·24)	0·15; 0·95 (0·88–1·51)	$4·80\times10^{-3}$	0·37; 0·96 (0·89–1·05)	$9·13\times10^{-4}$
FS	12	rs2238151	*ALDH2*	T	0·66	$1·03\times10^{-6}$; 1·13 (1·08–1·19)	0·26; 1·03 (0·98–1·09)	$8·62\times10^{-4}$	0·22; 1·04 (0·98–1·11)	$3·98\times10^{-4}$
YS	7	rs12703165	*PRKAG2*	G	0·82	$5·63\times10^{-7}$; 1·20 (1·12–1·29)	0·49; 0·98 (0·93–1·04)	0·012	0·89; 1·00 (0·94–1·06)	$1·81\times10^{-3}$
YS	8	rs4875812	*ARHGEF10*	G	0·55	$1·40\times10^{-6}$; 1·16 (1·10–1·23)	0·87; 1·00 (0·97–1·03)	0·034	0·94; 1·00 (0·97–1·03)	0·024

Chr=chromosome. SNP=single nucleotide polymorphism. RA=risk allele. RAF=risk allele frequency. $p_{discovery}$=one-sided p value in discovery cohorts. $OR_{discovery}$=odds ratio in discovery cohorts. $p_{replication}$=one-sided p value in replication cohorts. $OR_{replication}$=odds ratio in replication cohorts. $p_{combined}$=one-sided p value in all cohorts combined. IS=all ischaemic stroke. CS=cardioembolic stroke. LVD=large-vessel disease. SVD=small-vessel disease. FS=female-only stroke. YS=young stroke.

Table 3: **Association signals for SNPs selecting for testing in the independent replication cohort by subtype**

as the sample was taken at recruitment to the study and therefore before the onset of stroke.

By contrast, the *HDAC9* and 9p21 associations were specific to large-vessel stroke, and not present with other stroke subtypes. An association with the 9p21 locus was first associated with myocardial infarction and coronary artery disease but has now been associated more widely with other arterial diseases such as aneurysms and ischaemic stroke.[10,24] *HDAC9* was recently identified in the WTCCC2 ischaemic stroke study as a novel association with ischaemic stroke,[14] having not previously been shown in GWAS analyses of ischaemic heart disease.

For the *PITX2*, *ZFHX3*, and *HDAC9* associations, we did a conditional analysis to establish whether the lead SNP that we had identified was sufficient to model all of the associations within that region, or whether other independent genetic variants were associated with disease. In every case, no significant association remained after controlling for the lead SNP, suggesting that all the signal in each region can be attributed to one risk haplotype.

A meta-analysis of prospective cohort studies reported an association between ischaemic stroke and a SNP in the 12p13 region, although this was not replicated in an independent study.[13] The underlying gene was suggested to be *NINJ2*.[12] This association was present in the METASTROKE discovery cohort, but this cohort contained the datasets in which the original association had been determined. When these datasets were excluded, there was no evidence of any associations.

In a Japanese population, a variant in *PRKCH* has been associated with small-artery disease, a stroke subtype that is particularly common in this ethnic group.[11] This association was confirmed in a prospective study with relatively few stroke endpoints, and also in a Chinese population.[25,26] Interestingly, an association was also suggested with cerebral haemorrhage, which shares some underlying pathological similarities with cerebral small-vessel disease causing lacunar infarction. The association has not yet been examined in other ancestral groups. The SNP is monomorphic in European populations and therefore we were unable to examine whether the association was present in our population. However, we assessed all SNPs at this chromosomal region and noted no evidence of any association in our population of European ancestry.

We identified associations at four loci that were near genome-wide significance in the discovery cohort and had not been associated with stroke in previous studies: SNPs in the *ERRF11* and *NAA25* (C12orf30) genes with all ischaemic stroke, a SNP in *ALKBH8* with large-vessel stroke, and rs13407662 on chromosome 2p16.2 in an intergenic region with small-vessel disease. We took these four forward, with an additional eight of the strongest associations that had not reached genome-wide significance, to replication in an independent sample. None of the associations replicated. Our replication sample contained a cohort of patients of Pakistani ancestry, but, restriction of our analysis to individuals of European ancestry did not alter the results.

The same risk allele of SNP rs17696736 in the *NAA25* gene has previously been associated with type 1 diabetes in a large genome-wide association study.[27] Other SNPs in this 12q24 region have also been implicated in several of related phenotypes including microcirculation in vivo, platelet count, and blood pressure.[28–30] None of the other three associations near to genome-wide significance have previously been associated with cardiovascular or neurological disease.

	Chr	SNP	Candidate gene	RA	RAF	$p_{discovery}$; $OR_{discovery}$ (95% CI)	All replication samples		Replication in European descent individuals only	
							$p_{replication}$; $OR_{replication}$ (95% CI)	$p_{combined}$	$p_{replication}$; $OR_{replication}$ (95% CI)	$p_{combined}$
IS	1	rs225132	*ERRF11*	T	0·82	$6·27\times10^{-8}$; 1·12 (1·07–1·17)	0·16; 0·97 (0·92–1·01)	$1·65\times10^{-3}$	0·11; 0·96 (0·92–1·01)	$1·91\times10^{-3}$
IS	12	rs17696736	*NAA25 (C12orf30)*	G	0·42	$5·97\times10^{-8}$; 1·10 (1·06–1·14)	0·59; 1·01 (0·97–1·05)	$1·92\times10^{-5}$	0·60; 1·01 (0·97–1·05)	$1·69\times10^{-5}$
IS	3	rs16851055	*SPSB4*	G	0·81	$6·34\times10^{-7}$; 1·12 (1·07–1·17)	0·20; 1·03 (0·98–1·08)	$6·23\times10^{-4}$	0·25; 1·03 (0·98–1·08)	$7·76\times10^{-4}$
CS	3	rs6763538	*OXNAD1*	T	0·04	$2·89\times10^{-7}$; 1·47 (1·27–1·69)	0·69; 1·04 (0·87–1·24)	$2·68\times10^{-4}$	0·59; 1·05 (0·88–1·25)	$1·36\times10^{-4}$
LVD	11	rs7937106	*ALKBH8*	C	0·16	$5·85\times10^{-8}$; 1·68 (1·40–2·03)	0·66; 1·04 (0·87–1·25)	$3·93\times10^{-4}$	0·65; 1·05 (0·85–1·31)	$1·42\times10^{-4}$
LVD	6	rs556621	..	T	0·33	$4·63\times10^{-7}$; 1·20 (1·12–1·28)	0·46; 1·03 (0·96–1·10)	$5·33\times10^{-4}$	0·37; 1·03 (0·96–1·11)	$2·43\times10^{-4}$
SVD	18	rs7407640	*AFG3L2*	A	0·21	$2·20\times10^{-6}$; 1·23 (1·13–1·34)	0·99; 1·00 (0·91–1·10)	$4·54\times10^{-4}$	0·57; 0·97 (0·88–1·07)	$1·16\times10^{-3}$
SVD	2	rs13407662	..	T	0·04	$5·18\times10^{-8}$; 1·95 (1·53–2·48)	0·28; 1·16 (0·89–1·51)	$1·97\times10^{-4}$	0·36; 1·14 (0·86–1·53)	$1·88\times10^{-4}$
FS	3	rs7432308	..	T	0·15	$1·63\times10^{-6}$ 1·16 (1·09–1·24)	0·15; 0·95 (0·88–1·51)	$4·80\times10^{-3}$	0·37; 0·96 (0·89–1·05)	$9·13\times10^{-4}$
FS	12	rs2238151	*ALDH2*	T	0·66	$1·03\times10^{-6}$; 1·13 (1·08–1·19)	0·26; 1·03 (0·98–1·09)	$8·62\times10^{-4}$	0·22; 1·04 (0·98–1·11)	$3·98\times10^{-4}$
YS	7	rs12703165	*PRKAG2*	G	0·82	$5·63\times10^{-7}$; 1·20 (1·12–1·29)	0·49; 0·98 (0·93–1·04)	0·012	0·89; 1·00 (0·94–1·06)	$1·81\times10^{-3}$
YS	8	rs4875812	*ARHGEF10*	G	0·55	$1·40\times10^{-6}$; 1·16 (1·10–1·23)	0·87; 1·00 (0·97–1·03)	0·034	0·94; 1·00 (0·97–1·03)	0·024

Chr=chromosome. SNP=single nucleotide polymorphism. RA=risk allele. RAF=risk allele frequency. $p_{discovery}$=one-sided p value in discovery cohorts. $OR_{discovery}$=odds ratio in discovery cohorts. $p_{replication}$=one-sided p value in replication cohorts. $OR_{replication}$=odds ratio in replication cohorts. $p_{combined}$=one-sided p value in all cohorts combined. IS=all ischaemic stroke. CS=cardioembolic stroke. LVD=large-vessel disease. SVD=small-vessel disease. FS=female-only stroke. YS=young stroke.

Table 3: **Association signals for SNPs selecting for testing in the independent replication cohort by subtype**

as the sample was taken at recruitment to the study and therefore before the onset of stroke.

By contrast, the *HDAC9* and 9p21 associations were specific to large-vessel stroke, and not present with other stroke subtypes. An association with the 9p21 locus was first associated with myocardial infarction and coronary artery disease but has now been associated more widely with other arterial diseases such as aneurysms and ischaemic stroke.[10,24] *HDAC9* was recently identified in the WTCCC2 ischaemic stroke study as a novel association with ischaemic stroke,[14] having not previously been shown in GWAS analyses of ischaemic heart disease.

For the *PITX2*, *ZFHX3*, and *HDAC9* associations, we did a conditional analysis to establish whether the lead SNP that we had identified was sufficient to model all of the associations within that region, or whether other independent genetic variants were associated with disease. In every case, no significant association remained after controlling for the lead SNP, suggesting that all the signal in each region can be attributed to one risk haplotype.

A meta-analysis of prospective cohort studies reported an association between ischaemic stroke and a SNP in the 12p13 region, although this was not replicated in an independent study.[13] The underlying gene was suggested to be *NINJ2*.[12] This association was present in the METASTROKE discovery cohort, but this cohort contained the datasets in which the original association had been determined. When these datasets were excluded, there was no evidence of any associations.

In a Japanese population, a variant in *PRKCH* has been associated with small-artery disease, a stroke subtype that is particularly common in this ethnic group.[11] This association was confirmed in a prospective study with relatively few stroke endpoints, and also in a Chinese population.[25,26] Interestingly, an association was also suggested with cerebral haemorrhage, which shares some underlying pathological similarities with cerebral small-vessel disease causing lacunar infarction. The association has not yet been examined in other ancestral groups. The SNP is monomorphic in European populations and therefore we were unable to examine whether the association was present in our population. However, we assessed all SNPs at this chromosomal region and noted no evidence of any association in our population of European ancestry.

We identified associations at four loci that were near genome-wide significance in the discovery cohort and had not been associated with stroke in previous studies: SNPs in the *ERRF11* and *NAA25* (C12orf30) genes with all ischaemic stroke, a SNP in *ALKBH8* with large-vessel stroke, and rs13407662 on chromosome 2p16.2 in an intergenic region with small-vessel disease. We took these four forward, with an additional eight of the strongest associations that had not reached genome-wide significance, to replication in an independent sample. None of the associations replicated. Our replication sample contained a cohort of patients of Pakistani ancestry, but, restriction of our analysis to individuals of European ancestry did not alter the results.

The same risk allele of SNP rs17696736 in the *NAA25* gene has previously been associated with type 1 diabetes in a large genome-wide association study.[27] Other SNPs in this 12q24 region have also been implicated in several of related phenotypes including microcirculation in vivo, platelet count, and blood pressure.[28–30] None of the other three associations near to genome-wide significance have previously been associated with cardiovascular or neurological disease.

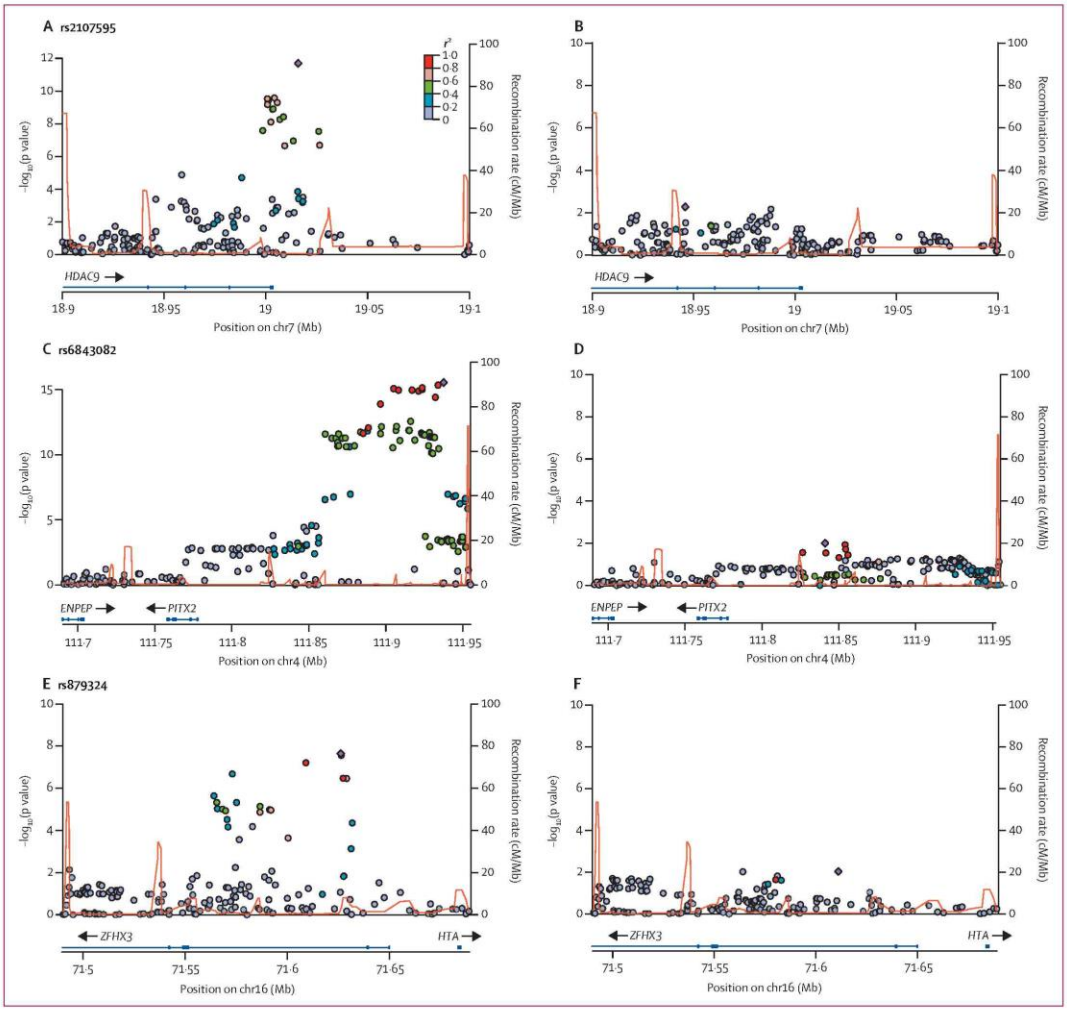

Figure 3: **Plots of conditional analysis regions before and after conditioning on lead SNP**

SNP=single nucleotide polymorphism. Plots of association signals around loci investigated in conditional analyses in subtypes in which they were discovered for the meta-analysed discovery samples. SNPs are coloured on the basis of their correlation (r^2) with the labelled top SNP, which has the smallest p value in the region. The fine-scale recombination rates estimated from HapMap data are marked in red, with genes marked below by horizontal blue lines. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. (A,C,E) Regions from discovery meta-analyses. (B,D,F) Same regions as A,C,E after conditioning on the lead SNP from the region.

Our inability to replicate any of the novel associations we identified in the discovery phase could be explained by various factors. All non-imputed SNPs in all cohorts were checked for Hardy-Weinberg equilibrium and standard quality control measures were done, including checking for sex mismatch on the basis of three genotypic markers, but we cannot rule out confounding by other means. For example, many of the 12 replication cohorts only directly genotyped the 12 replication SNPs. First, this type of analysis provides no means of adjustment for ancestry-informative principal components, which could lead to results being adversely affected by population

Panel: Research in context

Systematic review

As part of the International Stroke Genetics Consortium we had access to several genome-wide association datasets for ischaemic stroke, including both published and unpublished studies. To identify other studies, we searched PubMed on July 30, 2012, for published genome-wide association studies in ischaemic stroke with the terms "ischaemic stroke" and "genome wide association". The search returned studies already included by the consortium members. No further studies with ischaemic stroke as a primary endpoint were identified.

Interpretation

This is the largest analysis of genetic data for ischaemic stroke. This study provides evidence that common genetic variation has a role in the pathogenesis of ischaemic stroke. The genetic associations identified so far are with specific stroke subtypes, suggesting that the different subtypes of ischaemic stroke have different risk factor profiles and pathophysiological mechanisms, with potential implications for all areas of stroke research.

structure. Second, our strategy of attempting replication with one SNP from each region might not have been optimum. In regions such as the 12q24 locus, where the linkage disequilibrium patterns are complex, attempting replication in multiple SNPs might have proved more fruitful. Furthermore, one SNP (rs13407662) associated with small-vessel disease in the discovery phase failed genotyping in more than half of the replication cohorts. Genotyping multiple SNPs at this locus might have avoided this issue. We also cannot rule out confounding because of other environmental factors or phenotypic heterogeneity. Although phenotyping was done using the TOAST classification system, interpretation of exact classification criteria and definitions can differ across countries and studies, which becomes more of an issue when there are many smaller cohorts, such as in the replication phase of this study. Varying cohort study designs might also increase heterogeneity in large-scale meta-analyses.

Our results show that although genetic variants can be detected with ischaemic stroke, all associations we were able to confirm were specific to a stroke subtype. This finding has two implications. First, to maximise success of genetic studies in ischaemic stroke, detailed stroke subtyping is needed. Second, it implies that different pathophysiological mechanisms are associated with different stroke subtypes and, therefore, drug treatments might have different effects in different stroke subtypes. Most trials of secondary prevention in stroke have included all strokes, with limited stroke subtyping, and further studies with the detailed subtyping would be required to show different pharmacological profiles.

METASTROKE brings together GWAS data from most groups working in the area of stroke genetics worldwide. This paper describes the details of every population and represents the first analysis of the datasets. Various additional GWAS studies in stroke are currently taking place or have recently been completed, including a recently published GWAS in an Australian population, which confirmed an association at a 6p21.1 locus with large-artery atherosclerotic stroke.[31] The addition of these data might lead to identification of further novel associations with ischaemic stroke.

Affiliations

Stroke and Dementia Research Centre, St George's University of London, London, UK (M Traylor MSc, S Bevan PhD, H S Markus DM); Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK (M Farrall FRCPath); Department of Cardiovascular Medicine, University of Oxford, Oxford, UK (M Farrall, A Helgadottir MD); Centrw for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, and Center for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, NSW, Australia (E G Holliday PhD); Division of Clinical Neurosciences and Insititute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK (C Sudlow FRCP); Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, UK (J C Hopewell PhD, R Clarke FRCP); University of Maryland School of Medicine, Department of Medicine, Baltimore, MD, USA (Y-C Cheng PhD, B D Mitchell PhD); University of Texas Health Science Center at Houston, Houston, TX, USA (M Fornage PhD); Department of Epidemiology (M A Ikram MD, A Hofman MD), and Department of Neurology and Department of Radiology (M A Ikram), Erasmus MC University Medical Center, Rotterdam, Netherlands; Netherlands Consortium for Healthy Ageing, Leiden, Netherlands (M A Ikram, A Hofman); Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität, and Munich Cluster for Systems Neurology (SyNergy), Munich, Germany (R Malik PhD, A Gschwendtner MD, M Dichgans MD); deCODE Genetics, Reykjavik, Iceland (U Thorsteinsdottir PhD, A Helgadottir, G Thorleifsson PhD, S Gretarsdottir PhD, K Stefansson MD); Faculty of Medicine, University of Iceland, Reykjavik, Iceland (U Thorsteinsdottir, A Helgadottir, K Stefansson MD); Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA (M A Nalls PhD); Department of Neurology (W T Longstreth MD), Department of Epidemiology (W T Longstreth), and Cardiovascular Health Research Unit, Department of Medicine (K L Wiggins MS, J C Bis PhD), University of Washington, Seattle, WA, USA; Imperial College Cerebrovascular Research Unit (ICCRU), Imperial College London, London, UK (S Yadav MSc, M S Khan MSc, P Sharma PhD); Department of Cereberovascular Disease, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Neurologico Carlo Besta, Milan, Italy (E A Parati MD, G B Boncoraglio MD); Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA (A L DeStefano PhD); Department of Neurology, University of Virginia, Charlottesville, VA, USA (B B Worrall MD); Department of Public Health Science, University of Virginia, Charlottesville, VA, USA (B B Worrall, W-M Chen PhD); Department of Neurology, Veterans Affairs Medical Center, Baltimore, MA, USA, and Department of Neurology, University of Maryland School of Medicine, MA, USA (S J Kittner MD); Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA (A P Reiner MD, C Carty PhD); Department of Neurology and Neurosurgery, Utrecht Stroke Center, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, Netherlands (S Achterberg PhD, A Algra MD); Neurovascular Research Laboratory, Neurology and Medicine Departments, Universitat Autònoma de Barcelona and Institute of Research Vall d'Hebrón Hospital, Barcelona, Spain (I Fernandez-Cadenas PhD, J Montaner MD); Laboratory of Experimental Neurology, Brussels, Belgium (S Abboud MD, M Pandolfo MD); Department of Neurology, Division of Neurogeriatrics, Medical University Graz, Graz, Austria (R Schmidt MD); Institute of Cardiovascular and Medical Sciences,

University of Glasgow, Glasgow, UK (M Walters MD, P Higgins MRCP); Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA (W-M Chen, M Sale PhD); Department of Neurology, University of Münster, Münster, Germany (E B Ringelstein MD); National University of Ireland Galway, Galway, Ireland (M O'Donnell MD); Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK (W Kee Ho PhD, D Saleheen MD); Department of Neurology, Jagiellonian University, Krakow, Poland (J Pera MD, A Slowik MD); Laboratory of Neurobiology, Vesalius Research Center, VIB, Leuven, Belgium (R Lemmens MD, V Thijs MD); Experimental Neurology and Leuven Research Institute for Neurodegenerative Diseases (LIND), University of Leuven (KU Leuven), and Department of Neurology, University Hospital Leuven, Leuven, Belgium (R Lemmens, V Thijs); Department of Clinical Sciences Lund, Neurology, Lund University, and Department of Neurology, Skåne University Hospital, Lund, Sweden (B Norrving MD, H Delavaran MD, A Lindgren MD); Department of Clinical Biochemistry and The Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, and Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark (M Benn MD, B G Nordestgaard MD); Division of Cardiovascular Medicine, Department of Internal Medicine, University of Virginia, Charlottesville, VA, USA (M Sale); Institute for Experimental Medicine, University of Kiel, Germany (G Kuhlenbäumer MD); Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK (A S F Doney PhD, C N A Palmer PhD, N R van Zuydam MSc); Departamento Promoção da Saúde e Doenças Crónicas, Instituto Nacional de Saúde Dr Ricardo Jorge, Lisbon, Portugal (A M Vicente PhD); Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands (A Algra); Department of Psychology, and Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK (G Davies PhD, I Deary PhD); Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal (S A Oliveira PhD); Institute of Molecular Biology and Biochemistry, Medical University Graz, Graz, Austria (H Schmidt MD); Department of Medical Genetics and Department of Epidemiology, University Medical Centre Utrecht, Utrecht, Netherlands (P I W de Bakker PhD); Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA (P I W de Bakker, J Rosand MD); Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA (P I W de Bakker); Department of Neurology, Karolinska Institutet at Karolinska University Hospital, Huddinge, Sweden (K Kostulas MD); Serviço de Neurologia, Centro de Estudos Egas Moniz, Hospital de Santa Maria, Lisbon, Portugal (J M Ferro MD); Landspitali, University Hospital, Reykjavik, Iceland (E Valdimarsson MD); The Copenhagen City Heart Study, Bispebjerg Hospital, Copenhagen University Hospital, Copenhagen, Denmark (B G Nordestgaard); Centre for Non-Communicable Diseases, Karachi, Pakistan, and Department of Medicine, University of Pennsylvania, PA, USA (D Saleheen); Department of Pathology & Molecular Medicine and Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, ON, Canada (G Paré MD); Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany (K Berger MD); University of Mississippi Medical Center, Jackson, MS, USA (T H Mosley PhD, J Rosand); Department of Neurology, Massachusetts General Hospital, Boston, MA, USA (K Furie MD); Centre for Translational Neuroscience and Mental Health Research, University of Newcastle, and Hunter Medical Research Institute, New Lambton, NSW, Australia (C Levi MD); Department of Neurology, Boston University School of Medicine, Boston, MA, USA (S Seshadri MD); Department of Epidemiology, Department of Medicine, and Department of Health Services, University of Washington, and Group Health Research Institute, Group Health Seattle, WA, USA (B M Psaty); Stroke Prevention Research Unit, Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, UK (P M Rothwell FMedSci); Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA (J Rosand); and Department of Neurology, Mayo Clinic, Jacksonville, FL, USA (J F Meschia MD)

For more on the **Cardiovascular Health Study** see http://www.chs-nhlbi.org/pi.htm

Contributors

HM, MD, and MFa designed the experiment. MT drew the figures. MFa, MT, SB, and RM did the meta-analysis and subsequent replication statistical analysis. MT, MFa, EGH, CS, JCH, Y-CC, MFo, FAI, RM, SV, UT, MAN, WTL, KLW, SY, EAP, ALD, KS, BBW, SJK, MSK, AH, THM, BDM, KF, RC, CL, SS, AG, GBB, PS, JCB, BMP, PMR, JR, JFM, SG, MD, and HSM were responsible for the collection, phenotyping, or analysis of the discovery cohorts. Replication samples or replication data were provided by AR, AH, SAc, IF-C, SAb, RS, MW, W-MC, EBR, MO, WKH, JP, RL, BN, PH, MB, MS, GK, ASFD, AMV, HD, AA, GD, SAO, CNAP, ID, HS, MP, JM, CC, PIWD, KK, JMF, NRV, BGN, AL, VT, AS, DS, GP, KB, GT, and CS. SB coordinated wet lab replication genotyping. MT, HM, and MFa wrote the first draft of the report. All authors reviewed and commented on the report.

Conflicts of interest

All authors affiliated with deCODE are employees of deCODE, a biotechnology company. Some deCODE employees own stock options in deCODE. The other authors declare that they have no conflicts of interest.

Acknowledgments

Atherosclerosis Risk in Communities Study (ARIC) is a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; National Institutes of Health (NIH) contract HHSN268200625226C; and NHLBI contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, and grants R01-HL087641, U01 HL096917 (T H Mosley), and R01-HL093029 (M Fornage). Infrastructure was partly supported by Grant Number UL1RR025005, a component of the NIH and NIH Roadmap for Medical Research. ARIC analyses performed as part of the METASTROKE project were supported by grant HL-093029 to M Fornage. *Australian Stroke Genetics Collaboration (ASGC)* Australian population control data were derived from the Hunter Community Study. We also thank the University of Newcastle for funding and the men and women of the Hunter region who participated in this study. This research was funded by grants from the Australian National and Medical Health Research Council (NHMRC Project Grant ID: 569257), the Australian National Heart Foundation (NHF Project Grant ID: G 04S 1623), the University of Newcastle, the Gladys M Brawn Fellowship scheme, and the Vincent Fairfax Family Foundation in Australia. Elizabeth G Holliday is supported by the Australian NHMRC Fellowship scheme. *Bio-Repository of DNA in Stroke (BRAINS)* is partly funded by a Senior Fellowship from the Department of Health (UK) to P Sharma, the Henry Smith Charity and the UK-India Education Research Institutive (UKIERI) from the British Council. *Cardiovascular Health Study (CHS)* research was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, HHSN268201200036C and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from the National Institute of Neurological Diseases and Stroke (NINDS). Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the National Institute on Aging (NIA). DNA handling and genotyping was supported in part by National Center of Advancing Translational Technologies CTSI grant ULTTR000124 and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. *deCODE Genetics* Work performed at deCODE was funded in part through a grant from the European Community's Seventh Framework Programme (FP7/2007-2013), the ENGAGE project grant agreement HEALTH-F4-2007-201413. *Framingham Heart Study (FHS)* This work was supported by the dedication of the Framingham Heart Study participants, the NHLBI's Framingham Heart Study (Contract Nos. N01-HC-25195 and N02-HL-6-4278), and by grants from the NINDS (NS17950), the NHLBI (H193029), and the NIA (AG033193). *Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA* was supported by the NIH Genes,

Environment and Health Initiative (GEI) Grant U01 HG004436, as part of the GENEVA consortium under GEI, with additional support provided by the Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488), and the Office of Research and Development, Medical Research Service, and the Baltimore Geriatrics Research, Education, and Clinical Center of the Department of Veterans Affairs. Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the NIH to the Johns Hopkins University (contract number HHSN268200782096C). Assistance with data cleaning was provided by the GENEVA Coordinating Center (U01 HG 004446; PI Bruce S Weir). Study recruitment and assembly of datasets were supported by a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and Prevention and by grants from NINDS and the NIH Office of Research on Women's Health (R01 NS45012, U01 NS069208-01). *Heart Protection Study (HPS)* (ISRCTN48489393) was supported by the UK Medical Research Council (MRC), British Heart Foundation, Merck and Co (manufacturers of simvastatin), and Roche Vitamins Ltd (manufacturers of vitamins). Genotyping was supported by a grant to Oxford University and CNG from Merck and Co. Jemma C Hopewell acknowledges support from the British Heart Foundation Centre of Research Excellence, Oxford (RE/08/004). *Heart and Vascular Health Study (HVH)* research reported in this article was funded by NHLBI grants R01 HL085251 and R01 HL073410. *Ischemic Stroke Genetics Study (ISGS)/ Siblings With Ischemic Stroke Study (SWISS)* was supported in part by the Intramural Research Program of the NIA, NIH project Z01 AG-000954-06. ISGS/ SWISS used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds), human subjects protocol numbers 2003-081 and 2004-147. ISGS/SWISS used stroke-free participants from the Baltimore Longitudinal Study of Aging (BLSA) as controls. The inclusion of BLSA samples was supported in part by the Intramural Research Program of the NIA, NIH project Z01 AG-000015-50, human subjects protocol number 2003-078.The ISGS study was funded by NIH-NINDS Grant R01 NS-42733 (J F Meschia). The SWISS study was funded by NIH-NINDS Grant R01 NS-39987 (J F Meschia). This study used the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (http://biowulf.nih.gov). *MGH Genes Affecting Stroke Risk and Outcome Study (MGH-GASROS)* was supported by NINDS (U01 NS069208), the American Heart Association/Bugher Foundation Centers for Stroke Prevention Research 0775010N, the NIH and NHLBI's STAMPEED genomics research program (R01 HL087676), and a grant from the National Center for Research Resources. The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278 from the National Center for Research resources. *Milano - Besta Stroke Register* Collection and genotyping of the Milan cases within CEDIR were supported by Annual Research Funding of the Italian Ministry of Health (Grant Numbers: RC 2007/LR6, RC 2008/LR6; RC 2009/LR8; RC 2010/LR8). FP6 LSHM-CT-2007-037273 for the PROCARDIS control samples. *Rotterdam Study* was supported by the Netherlands Organization of Scientific Research (175.010.2005.011), the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Ageing (050-060-810), the Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and Development, the Research Institute for Diseases in the Elderly, the Ministry of Education, Culture, and Science, the Ministry for Health, Welfare, and Sports, the European Commission, and the Municipality of Rotterdam to the Rotterdam Study. Further funding was obtained from the Netherlands Heart Foundation (Nederlandse Hartstichting) 2009B102. *Wellcome Trust Case-Control Consortium 2 (WTCCC2)* was principally funded by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). The Stroke Association provided additional support for collection of some of the St George's, London cases. The Oxford cases were collected as part of the Oxford Vascular Study which is funded by the MRC, Stroke Association, Dunhill Medical Trust, National Institute of Health Research (NIHR) and the NIHR Biomedical Research Centre, Oxford. The Edinburgh Stroke Study was supported by the Wellcome Trust (clinician scientist award to C Sudlow), and the Binks Trust. Sample processing occurred in the Genetics Core Laboratory of the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. Much of the neuroimaging occurred in the Scottish Funding Council Brain Imaging Research Centre (www.sbirc.ed.ac.uk), Division of Clinical Neurosciences, University of Edinburgh, a core area of the Wellcome Trust Clinical Research Facility and part of the SINAPSE (Scottish Imaging Network—A Platform for Scientific Excellence) collaboration (www.sinapse.ac.uk), funded by the Scottish Funding Council and the Chief Scientist Office. Collection of the Munich cases and data analysis was supported by the Vascular Dementia Research Foundation. M Farrall and A Helgadottir acknowledge support from the BHF Centre of Research Excellence in Oxford and the Wellcome Trust core award (090532/Z/09/Z). *Barcelona* The Neurovascular Research Laboratory takes part in the International Stroke Genetics Consortium (ISGC), the Spanish Stroke Genetics Consortium (www.genestroke.com), and the Cooperative Neurovascular Research RENEVAS (RD06/0026/0010). This study was funded by a grant of the Spanish government (PI10/01212.). The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreements #201024 and #202213 (European Stroke Network). *Belgium Stroke Study (BSS)* was supported by Erasme Funds. *Edinburgh Stroke Study (ESS)* (which contributed discovery cases as part of WTCCC2 and additional replication cases) was supported as described above. *Lothian Birth Cohort 1936* was supported in part by Research into Aging, Help the Aged (Sidney De Haan Award and The Disconnected Mind Major Gift Campaign), MRC, and UK Biotechnology and Biological Sciences Research Council (BBSRC). Lothian Birth Cohort 1936 was also supported by a programme grant from Research Into Ageing and continues with programme grants from Help the Aged/Research Into Ageing (Disconnected Mind). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the BBSRC, Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC), and MRC is gratefully acknowledged. Genotyping of the LBC1936 was funded by the BBSRC. *Glasgow* The work was supported by NHS Greater Glasgow Endowment funds. *Genetics of Diabetes Audit and Research in Tayside Study (Go-Darts)* N R van Zuydam is supported by PhD funding from IMI SUMMIT study, under the EU Framework Programme 7 funding stream. The Wellcome Trust provides support for Wellcome Trust United Kingdom Type 2 Diabetes Case Control Collection (Go-DARTS) and the Scottish Health Informatics Programme. *Graz Stroke Study* Genetic studies of the Austrian Stroke Prevention Study are supported by the Austrian Science Fund (P20545). *Interstroke* has received unrestricted grants from the Canadian Institutes of Health Research, Heart and Stroke Foundation of Canada, Canadian Stroke Network, Pfizer Cardiovascular Award, Merck, AstraZeneca, and Boehringer Ingelheim. The study was facilitated by CANNeCTIN network. Funding for genotyping was provided by the Heart and Stroke Foundation of Ontario, the Canadian Stroke Network, and by an unrestricted grant from Boehringer-Ingelheim. *Leuven Stroke Study* is funded by personal research funds from the Department of Neurology, University Hospital Leuven, Leuven, Belgium. V Thijs and R Lemmens are supported by Fundamental Clinical Investigatorships from FWO Flanders. *Lund Stroke Register* (LSR) was supported by the Swedish Research Council (K2007-61X-20378-01-3, K2010-61X-20378-04-3), Region Skåne, the Freemasons Lodge of Instruction EOS in Lund, King Gustav V and Queen Victoria's Foundation, Lund University, and the Swedish Stroke Association. DNA extraction and preparation for LSR was performed by the SWEGENE Resource Center for Profiling Polygenic Disease (Skåne University Hospital, Malmö, Sweden). *Munster (Westphalian Stroke Cases and Controls from the Dortmund Health Study, Germany)* Case ascertainment in the Westphalian Stroke Register was part of the German Competence Net Stroke, supported by the German Federal Ministry of Education and Research (01GI9909/3). Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine University of Muenster. The collection of sociodemographic and clinical data in the Dortmund Health Study was supported by the German Migraine and Headache Society (DMKG) and by unrestricted

grants of equal share from Almirall, Astra Zeneca, Berlin Chemie, Boehringer, Boots Health Care, Glaxo-Smith-Kline, Janssen Cilag, McNeil Pharma, MSD Sharp & Dohme, and Pfizer to the University of Muenster. *Portugal* SAO, JMF, and AMV are deeply grateful to all study participants, to the genotyping unit at the Instituto Gulbenkian de Ciência, and to the Portuguese study neurologists and nurses for their contributions. This work was supported by the PTDC/SAU-GMG/64426/2006 grant, a Ciência 2008 contract (SAO) and doctoral fellowships from the Portuguese Fundação para a Ciência e a Tecnologia. *Poland: Krakow* The study was supported by the grant from the Jagiellonian University, Krakow Poland: K/ZDS/002848. *SMART-study the Netherlands* Genotyping in the SMART Study was made possible, in part, by a Complementation Grant to P I W de Bakker from the Biobanking and Biomolecular Research Infrastructure in the Netherlands (BBMRI-NL). S Achterberg is working in part on a grant from the Netherlands Heart Foundation, No. 2005B031. *VISP* The GWAS component of the VISP study was supported by the United States National Human Genome Research Institute (NHGRI), Grant U01 HG005160 (PI Michèle Sale & Bradford Worrall), as part of the Genomics and Randomized Trials Network (GARNET). Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the NIH to the Johns Hopkins University. Assistance with data cleaning was provided by the GARNET Coordinating Center (U01 HG005157; PI Bruce S Weir). Study recruitment and collection of datasets for the VISP clinical trial were supported by an investigator-initiated research grant (R01 NS34447; PI James Toole) from the United States Public Health Service, NINDS, Bethesda, Maryland. Control data obtained through the database of genotypes and phenotypes (dbGAP) maintained and supported by the United States National Center for Biotechnology Information, US National Library of Medicine. *WHI* Funding support for WHI-GARNET was provided through the NHGRI GARNET (Grant Number U01 HG005152). Assistance with phenotype harmonisation and genotype cleaning, as well as with general study coordination, was provided by the GARNET Coordinating Center (U01 HG005157). Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH Genes, Environment, and Health Initiative (GEI; U01 HG004424).

References

1 Department of Health. Reducing brain damage: faster access to better stroke care. London: National Audit Office, 2005.
2 Sacco RL, Ellenberg JH, Mohr JP, et al. Infarcts of undetermined cause: the NINCDS Stroke Data Bank. *Ann Neurol* 1989; **25:** 382–90.
3 Dichgans M. Genetics of ischaemic stroke. *Lancet Neurol* 2007; **6:** 149–61.
4 Dichgans M, Markus HS. Genetic association studies in stroke: methodological issues and proposed standard criteria. *Stroke* 2005; **36:** 2027–31.
5 Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; **461:** 747–53.
6 Markus HS. Stroke genetics. *Hum Mol Genet* 2011; **20:** R124–31.
7 Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke* 2003; **34:** 1364–69.
8 Gretarsdottir S, Thorleifsson G, Manolescu A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol* 2008; **64:** 402–09.
9 Gudbjartsson DF, Holm H, Gretarsdottir S, et al. A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet* 2009; **41:** 876–78.
10 Gschwendtner A, Bevan S, Cole JW, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol* 2009; **65:** 531–39.
11 Kubo M, Hata J, Ninomiya T, et al. A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. *Nat Genet* 2007; **39:** 212–17.
12 Ikram MA, Seshadri S, Bis JC, et al. Genomewide association studies of stroke. *N Engl J Med* 2009; **360:** 1718–28.
13 International Stroke Genetics Consortium, Wellcome Trust Case–Control Consortium 2. Failure to validate association between 12p13 variants and ischemic stroke. *N Engl J Med* 2010; **362:** 1547–50.
14 International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2), Bellenguez C, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet* 2012; **44:** 328–33.
15 Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010; **42:** 1118–25.
16 Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; **41:** 703–07.
17 International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* 2011; **377:** 641–49.
18 Adams HP Jr, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*; **24:** 35–41.
19 International HapMap Consortium, Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449:** 851–61.
20 International HapMap 3 Consortium, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467:** 52–58.
21 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467:** 1061–73.
22 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **27:** 2190–91.
23 Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005; **95:** 221–27.
24 Helgadottir A, Thorleifsson G, Magnusson KP, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet* 2008; **40:** 217–24.
25 Wu L, Shen Y, Liu X, et al. The 1425G/A SNP in PRKCH is associated with ischemic stroke and cerebral hemorrhage in a Chinese population. *Stroke* 2009; **40:** 2973–76.
26 Serizawa M, Nabika T, Ochiai Y, et al. Association between PRKCH gene polymorphisms and subcortical silent brain infarction. *Atherosclerosis* 2008; **199:** 340–45.
27 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447:** 661–78.
28 Ikram MK, Sim X, Jensen RA, et al. Four novel loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS Genet* 2010; **6:** e1001184.
29 Soranzo N, Spector TD, Mangino M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* 2009; **41:** 1182–90.
30 Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* 2009; **41:** 666–76.
31 Holliday EG, Maguire JM, Evans TJ, et al. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet* 2012; published online Sept 2. DOI:10.1038/ng.2397.

Appendix 10. Candidate Gene Association Study for Diabetic Retinopathy in Persons with Type 2 Diabetes: The Candidate Gene Association Resource (CARe)

Genetics

Candidate Gene Association Study for Diabetic Retinopathy in Persons with Type 2 Diabetes: The Candidate Gene Association Resource (CARe)

Lucia Sobrin,[1,2] Todd Green,[2,3,4] Xueling Sim,[5] Richard A. Jensen,[6] E. Shyong Tai,[7,8,9] Wan Ting Tay,[10] Jie Jin Wang,[11,12] Paul Mitchell,[11] Niina Sandholm,[13,14] Yiyuan Liu,[15] Kustaa Hietala,[14,16] Sudha K. Iyengar,[17] for the Family Investigation of Nephropathy and Diabetes-Eye Research Group,[18] Matthew Brooks,[19] Monika Buraczynska,[20] Natalie Van Zuydam,[15] Albert V. Smith,[21,22] Vilmundur Gudnason,[21,22] Alex S. F. Doney,[23] Andrew D. Morris,[23] Graham P. Leese,[23] Colin N. A. Palmer,[15] for the Wellcome Trust Case Control Consortium 2,[24] Anand Swaroop,[19] Herman A. Taylor, Jr,[25] James G. Wilson,[26] Alan Penman,[25,27] Ching J. Chen,[28] Per-Henrik Groop,[13,14] Seang-Mei Saw,[8] Tin Aung,[10,29] Barbara E. Klein,[30] Jerome I. Rotter,[31] David S. Siscovick,[6] Mary Frances Cotch,[32] Ronald Klein,[30] Mark J. Daly,[2,3,4] and Tien Y. Wong[10,12,29]

From the [1]Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts; the [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts; the [4]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts; the [5]Centre for Molecular Epidemiology and the Departments of [7]Medicine, [8]Epidemiology and Public Health, and [29]Ophthalmology, National University of Singapore, Singapore; the [6]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington; the [9]Duke-National University of Singapore Graduate Medical School, Singapore; the [10]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; the [11]Centre for Vision Research, University of Sydney, Sydney, New South Wales, Australia; the [12]Centre for Eye Research Australia, University of Melbourne, Melbourne, Victoria, Australia; the [13]Division of Nephrology, Department of Medicine, Helsinki University, Helsinki, Finland; the [14]Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Helsinki, Finland; the [15]Biomedical Research Institute, University of Dundee, Dundee, Scotland, United Kingdom; the [16]Department of Ophthalmology, Helsinki University Central Hospital, Helsinki, Finland; the [17]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio; the [18]Family Investigation of Nephropathy and Diabetes-Eye Research Group (members listed in the Supplementary Material, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental); the [19]Neurobiology Neurodegeneration and Repair Laboratory and the [32]Division of Epidemiology and Clinical Applications, Intramural Research Program, National Eye Institute, National Institutes of Health, Bethesda, Maryland; the [20]Laboratory for DNA Analysis and Molecular Diagnostics, Department of Nephrology, Medical University of Lublin, Lublin, Poland; the [21]Icelandic Heart Association, Kopavogur, Iceland; the [22]University of Iceland, Reykjavik, Iceland; the [23]Department of Medicine and Therapeutics, Population Pharmacogenetics Group, Ninewells Hospital and Medical School, Dundee, Scotland, United Kingdom; the [24]Wellcome Trust Case Control Consortium (members listed in the Supplementary Material, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental); the Departments of [25]Medicine, [26]Physiology and Biophysics, [27]Biostatistics, and [28]Ophthalmology, University of Mississippi Medical Center, Jackson, Mississippi; the [30]Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; and the [31]Medical Genetics Institute, Cedar-Sinai Medical Center, Los Angeles, California. [2]These authors contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Supported by National Eye Institute Grant K12-EY16335; a Research to Prevent Blindness Career Development Award; the Massachusetts Lions Eye Research Fund; the Sara Elizabeth O'Brien Trust; and Harvard Catalyst/The Harvard Clinical and Translational Science Center, National Institutes of Health Award Grant UL1 RR 025758 and financial contributions from Harvard University and its affiliated academic health care centers. The National Heart, Lung Blood Institute (NHLBI) Grant N01-HC-65226 has supported genotyping and has created a genotype–phenotype database with data and samples from nine cohorts as part of Candidate Gene Association Resource (CARe). Please see the Supplementary Material (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental), as well as http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx for detailed information about the grants supporting the four cohorts in this study: Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), Jackson Heart Study (JHS), and Multi-ethnic Study of Atherosclerosis (MESA). Please see the Supplementary Material (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental) for detailed information about the grants supporting the replication cohorts.

Submitted for publication March 5, 2011; revised June 30 and August 3, 2011; accepted August 5, 2011.

Disclosure: **L. Sobrin**, None; **T. Green**, None; **X. Sim**, None; **R.A. Jensen** , None; **E.S. Tai**, None; **W.T. Tay**, None; **J.J. Wang**, None; **P. Mitchell**, None; **N. Sandholm**, None; **Y. Liu**, None; **K. Hietala**, None; **S.K. Iyengar**, None; **M. Brooks**, None; **M. Buraczynska**, None; **N. Van Zuydam**, None; **A.V. Smith** , None; **V. Gudnason**, None; **A.S.F. Doney**, None; **A.D. Morris**, None; **G.P. Leese**, None; **C.N.A. Palmer**, None; **A. Swaroop**, None; **H.A. Taylor, Jr**, None; **J.G. Wilson**, None; **A. Penman**, None; **C.J. Chen**, None; **P.-H. Groop**, None; **S.-M. Saw**, None; **T. Aung**, None; **B.E. Klein**, None; **J.I. Rotter**, None; **D.S. Siscovick**, None; **M.F. Cotch**, None; **R. Klein**, None; **M.J. Daly**, None; **T.Y. Wong**, None

Corresponding author: Lucia Sobrin, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114; lucia_sobrin@meei.harvard.edu.

Investigative Ophthalmology & Visual Science, September 2011, Vol. 52, No. 10

Purpose. To investigate whether variants in cardiovascular candidate genes, some of which have been previously associated with type 2 diabetes (T2D), diabetic retinopathy (DR), and diabetic nephropathy (DN), are associated with DR in the Candidate gene Association Resource (CARe).

Methods. Persons with T2D who were enrolled in the study (n = 2691) had fundus photography and genotyping of single nucleotide polymorphisms (SNPs) in 2000 candidate genes. Two case definitions were investigated: Early Treatment Diabetic Retinopathy Study (ETDRS) grades ≥14 and ≥30. The χ^2 analyses for each CARe cohort were combined by Cochran-Mantel-Haenszel (CMH) pooling of odds ratios (ORs) and corrected for multiple hypothesis testing. Logistic regression was performed with adjustment for other DR risk factors. Results from replication in independent cohorts were analyzed with CMH meta-analysis methods.

Results. Among 39 genes previously associated with DR, DN, or T2D, three SNPs in P-selectin (*SELP*) were associated with DR. The strongest association was to rs6128 (OR = 0.43, P = 0.0001, after Bonferroni correction). These associations remained significant after adjustment for DR risk factors. Among other genes examined, several variants were associated with DR with significant P values, including rs6856425 tagging α-L-iduronidase (*IDUA*) ($P = 2.1 \times 10^{-5}$, after Bonferroni correction). However, replication in independent cohorts did not reveal study-wide significant effects. The P values after replication were 0.55 and 0.10 for rs6128 and rs6856425, respectively.

Conclusions. Genes associated with DN, T2D, and vascular diseases do not appear to be consistently associated with DR. A few genetic variants associated with DR, particularly those in *SELP* and near *IDUA*, should be investigated in additional DR cohorts. (*Invest Ophthalmol Vis Sci.* 2011;52:7593–7602) DOI: 10.1167/iovs.11-7510

Diabetic retinopathy (DR) is the leading cause of blindness in working-age Americans[1,2] and is increasing in prevalence as rates of type 2 diabetes (T2D) soar worldwide.[3,4] The frequency and severity of DR are heterogeneous within and across ethnic groups,[5] even with adjustment for risk factors such as duration of diabetes and glycemic control.[2,6,7] There are people who have a long duration of diabetes without DR and those who have severe DR despite relatively good glycemic control. For these reasons, genetic risk factors are thought to play a role in DR. Heritability has been estimated to be as high as 27% for DR and 52% for proliferative diabetic retinopathy (PDR).[8–10] However, genetic association studies for DR have been thus far limited mostly to studies of one or a modest number of candidate genes.[11,12] Most reported associations have not been consistently reproduced.[11,13,14]

In contrast to DR, genetic association studies for T2D have revealed many consistently associated genes. Genes that increase T2D risk may also predispose to development of retinopathy. In the case of diabetic nephropathy (DN), a *TCF7L2* variant increases the risk of developing DN beyond the risk of diabetes.[15] Because there is evidence that DR shares risk factors and pathophysiological mechanisms with DN and macrovascular diabetic complications,[6,16–21] genes associated with DN and atherosclerotic vascular disease may also be associated with DR.

The Candidate gene Association Resource (CARe) is a collaboration for association analyses of genotypes and cardiovascular disease phenotypes.[22] It comprises >40,000 participants from nine cohorts who have been genotyped for 49,320 single nucleotide polymorphisms (SNPs) from approximately 2,000 candidate genes postulated or known to increase risk of cardiovascular, metabolic, and inflammatory diseases.[23] It includes 2691 T2D participants with fundus photographs of multiple ethnicities. Thus, the CARe framework provides an opportunity to investigate genetic associations for DR with a candidate gene approach. CARe genotyped many genes previously associated with DR,[24,25] DN,[25–27] and T2D.[28–34] The first purpose of this study was to investigate whether these genes are also associated with the presence of DR in CARe. The second purpose was to determine whether the remaining genes included in the CARe genotyping platform, which were also chosen as potential cardiovascular disease genes, are associated with DR.

Methods

Study Population and Fundus Photography Procedures

Four CARe cohorts have fundus photographs of T2D participants: Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), Jackson Heart Study (JHS), and Multiethnic Study of Atherosclerosis (MESA).[35–38] T2D was defined according to the American Diabetes Association 2003 Criteria.[39] The fundus photography protocol for each cohort is described in Table 1.[40–42] In all studies, except for the JHS, fundus photographs were graded by masked readers at the University of Wisconsin Ocular Epidemiology Reading Center according to the modified Airlie House Classification system.[43] Fundus photographs for the JHS were graded by masked JHS ophthalmologist investigators according to the same criteria.

Definition of Diabetic Retinopathy

We examined two DR phenotypes. First we defined cases as participants with an Early Treatment Diabetic Retinopathy Study (ETDRS) grade ≥14 in the eye with the higher ETDRS grade or in the only eye photographed, depending on the study's protocol. These analyses were designed to detect associations with the presence of any DR. Our second phenotype defined cases as participants with ETDRS grade ≥30. The latter was intended to reduce misclassification of patients with minimal signs of DR, which may be seen even in persons without diabetes.[44–46] For all analyses, controls were defined as T2D participants with an ETDRS grade <14 (no DR).

Measurement of Other Variables

Data on DR risk factors were obtained from the study examination at which fundus photography was performed. These included duration of diabetes, fasting blood glucose, systolic and diastolic blood pressures, and fasting total cholesterol.[47–50] The procedures for measuring these variables are described in online documentation (www.cscc.unc.edu/aric/, www.chs-nhlbi.org, and www.mesa-nhlbi.org, provided by the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD; and www.jsums.edu/jhs, JHS, Jackson State University, Jackson, MS). Some participants were unaware of a diabetes diagnosis and received the diagnosis based on their laboratory values at the study visit at which they also had fundus photography. For these patients, the duration of diabetes was calculated by halving the number of years between their prior study visit (when they did not meet criteria for T2D) and the visit at which they met criteria. If data on a risk factor were not measured at the fundus photography visit, the information was obtained from the visit closest in time to the fundus photography visit.

Genotyping

CARe participant DNA samples were interrogated on a custom genotyping array (iSelect ITMAT-Broad-CARe [IBC] Chip; Illumina, San Diego, CA). Its design is described elsewhere.[23] SNP selection criteria and genotyping quality control (QC) procedures are explained in the Supplementary Methods (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental).[23]

TABLE 1. CARe Participants with T2D and DR Grading by Cohort and Ethnicity

Cohort/ Population	Eyes Photographed per Participant (*n*)	Fields Photographed per Eye (*n*)	Size of Each Field Photographed (deg)	Participants with T2D, DR Grading and IBC Chip Genotyping (*n*)	Participants with ETDRS Grade <14 (*n*)	Participants with ETDRS Grade ≥14(*n*)	Participants with ETDRS Grade ≥30 (*n*)
ARIC							
EA	One	One	45	885	732	153	91
AfrA				439	315	124	95
CHS							
EA	One	One	45	193	160	33	20
AfrA				54	35	19	15
JHS							
AfrA	Two	Seven	30	55	26	29	22
MESA							
EA	Two	Two	45	176	140	36	11
AfrA				275	176	99	57
AsA				79	54	25	14
HA				231	151	80	46
All							
EA				1254	1032	222	122
AfrA				823	552	271	189
AsA				79	54	25	14
HA				231	151	80	46

EA, European American; AfrA, African American; AsA, Asian American; HA, Hispanic American.

Statistical Analysis

We first investigated genes that have been previously associated with DR, DN, and T2D. For DR, we chose the genes that had the most robust evidence of association from a comprehensive review of the literature[24] and a subsequent strong association with the erythropoietin (*EPO*) gene promoter.[25] For DN, we chose genes that have shown nominal associations ($P < 0.05$) with DN or a related quantitative trait.[25–27] For T2D, we chose genes with SNPs that had met genome-wide significance.[28–31,33] Supplementary Table S1 (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental) lists the 39 genes included on the IBC chip that met these criteria. In the second phase, we investigated the remaining genes on the IBC chip which were primarily cardiovascular candidate genes.[23]

For genetic association testing, we used χ^2 analysis comparing European-American cases (participants with T2D and DR) to controls (participants with T2D and no DR) in each CARe cohort. Results were combined by Cochran-Mantel-Haenszel (CMH) pooling of the odds ratios (ORs).[51,52] This CMH method is a robust way of maintaining consistency with individual study ORs while estimating a single fixed-effects OR across all cohorts. We report correction for the multiple association tests performed with per gene Bonferroni correction and permutation testing. Per gene Bonferroni correction was performed because it is an intuitive, easily understood correction. This Bonferroni correction is not as conservative as the one that corrects for the total number of unique genetic loci tested, as would be represented by correcting for the total number of tag SNPs that are not in strong LD. For this reason, we also present the empiric permutation testing correction, which, although less intuitive, does account for this LD between SNPs. We defined statistical significance as $P < 0.05$ after per gene Bonferroni correction. We chose this threshold for the discovery phase of the experiment, although it is a less stringent threshold than one based on a correction that could completely account for LD between SNPs, to minimize type II error (false negatives) in this initial phase. We apply a more stringent threshold for the replication phase (see below), and the final decision of whether an SNP is truly associated is based on its final *P* value after replication. For SNPs that were statistically significant in the discovery phase, we performed haplotype analyses using the omnibus test and further examined the associations with logistic regression models that included other DR risk factors. Age and duration of diabetes were defined as continuous variables in years. Fasting glucose and total cholesterol were incorporated as continuous variables in milligrams per deciliter. Systolic and diastolic blood pressures were evaluated as continuous variables (in mm Hg). If a participant was taking antihypertension medication, 15 and 10 mm Hg were added to the systolic and diastolic blood pressure values, respectively.[53] Sex and study site were also incorporated. All statistical analyses were performed in PLINK.[54]

Replication

Top significant findings were pursued in the non-European American populations in CARe and in independent Caucasian cohorts with genome-wide genotyping results: Age, Gene/Environment Susceptibility (AGES) study[55]; Blue Mountains Eye Study (BMES)[56]; Genetics of Diabetes Audit and Research Tayside Study (Go-DARTS)[57]; Finnish Diabetic Nephropathy (FinnDiane) Study[9]; Family Investigation of Nephropathy and Diabetes-Eye (FIND-Eye) Study[10]; Singapore Malay Eye Study (SiMES)[58]; and the Singapore Prospective Study Program (SP2).[59] The Medical University of Lublin T2D cohort[60] performed de novo genotyping for replication. The phenotyping protocols for these studies are described in the Supplementary Methods and Supplementary Table S2 (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). Because ETDRS grading was not used by all cohorts, phenotype data were harmonized into two categories that were analogous to ETDRS grade ≥14 and ETDRS grade ≥30. Meta-analysis of the discovery cohort and replication cohorts results was performed by CMH pooling of ORs. Statistical heterogeneity was assessed with Cochran's Q statistic.[61,62] The Q statistic calculates a

weighted sum of the square distances of the observed effects from the null hypothesis of equality of the effects. The weight for each study is the inverse of the variance of the effect estimator so that larger and more accurate studies are weighted more heavily. Statistical significance after replication was defined as a meta-analysis $P < 1 \times 10^{-6}$. This threshold was determined empirically on the basis of the number of genes tested, is analogous to the threshold for genome-wide significance of 5.0×10^{-8} for GWAS,[63] and corresponds approximately to a $P < 0.05$, after Bonferroni correction.

This research adhered to the tenets of the Declaration of Helsinki and was approved by the institutional review boards of the Massachusetts Eye and Ear Infirmary and the primary cohorts. Informed consent was obtained from all participants.

Results

Table 1 shows the number of T2D CARe participants by cohort and ethnicity. The prevalence of DR is similar among the cohorts, although it is higher in the JHS African-American population when compared to the other African-American populations. This discrepancy is probably secondary to the more precise phenotyping performed in the JHS with seven-field photography. Duration of diabetes and fasting glucose levels were not significantly different between the JHS cohort's and the other cohorts' African-American populations.

Table 2 shows the most significant associations for the analysis of the DR, DN, and T2D genes with any DR in European Americans. Only the associations to three SNPs in the P-selectin gene (*SELP*) were significant ($P < 0.05$, after Bonferroni correction). The three associated SNPs tagged the only associated haplotype (Supplementary Table S3, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). In logistic regression including other DR risk factors, the associations to rs6128, rs6133, and rs3917779 remained significant ($P = 0.026$, 0.022, and 0.026, respectively). The mean values for covariates are presented in Supplementary Table S4 (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental).

Table 3 shows the most significant associations from the analysis of the DR, DN, and T2D candidate genes with DR defined as an ETDRS grade ≥30 in European Americans. The three *SELP* SNPs, along with five SNPs in the fat mass and obesity-associated (*FTO*) gene, were significantly associated ($P < 0.05$, after Bonferroni correction). Variants rs12708942, rs9806929, and rs4783824 tagged the only associated haplotype (Supplementary Table S3, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). Because the effect of *FTO* on T2D risk is mediated by its effect on obesity, we executed a logistic regression model including age, sex, and body mass index (BMI). BMI was incorporated as a continuous variable (weight in kilograms divided by height in meters squared). Three *FTO* SNPs continued to be associated with DR ($P = 0.002$, 0.002, and 0.001 for rs9926180, rs7500562, and rs12149433, respectively). Of note, an *EPO* variant, rs551238, was associated with this more stringent definition of DR and had a $P < 0.05$ after permutation correction but not after Bonferroni correction.

We then examined the strength of these associations in the CARe non-European populations. For any DR, the associations with the *SELP* SNPs were not replicated in the African-, Hispanic-, or Asian-American populations (Table 4). When cases were defined as ETDRS grade ≥30, the SNPs in *SELP* and *FTO* were not associated, and in logistic regression models including age, sex, and BMI, no association with the *FTO* SNPs was detected (data not shown).

In the second phase of the analysis, we examined the remaining genes on the IBC chip in European Americans. The

Table 2. Top Association Results in European-American Samples with Cases Defined as ETDRS Grade ≥14 for Variants within Genes Previously Associated with DR, DN, or T2D

					ARIC				CHS				MESA				CMH Combined Analysis						P Correction	
					MAF				MAF				MAF				MAF							
SNP	Chr	Gene	Position	Minor Allele	Cases	Controls	OR	P	Cases	Controls	OR	P	Cases	Controls	OR	P	Cases	Controls	OR	L95	U95	P	Bonferroni	Permutation
rs6128	1	*SELP*	167829528	T	0.08	0.16	0.48	0.0007	0.06	0.18	0.30	0.02	0.08	0.20	0.36	0.02	0.08	0.17	0.43	0.31	0.62	3.1×10^{-6}	0.0001	0.0001
rs6133	1	*SELP*	167831970	A	0.05	0.12	0.39	0.0003	0.03	0.13	0.22	0.02	0.07	0.13	0.50	0.16	0.05	0.12	0.38	0.25	0.59	1.1×10^{-5}	0.0004	0.0001
rs3917779	1	*SELP*	167837472	A	0.05	0.12	0.40	0.0004	0.03	0.12	0.23	0.03	0.07	0.13	0.50	0.16	0.05	0.12	0.39	0.26	0.60	1.6×10^{-5}	0.0006	0.0001
rs12262390	10	*HHEX*	94436103	C	0.12	0.09	1.48	0.05	0.15	0.10	1.61	0.22	0.18	0.09	2.27	0.03	0.14	0.09	1.61	1.18	2.20	0.002	0.10	0.004
rs9356754	6	*CDKAL1*	20916721	G	0.51	0.43	1.37	0.01	0.53	0.44	1.43	0.18	0.40	0.38	1.09	0.74	0.49	0.42	1.32	1.08	1.63	0.007	0.29	0.008
rs9465904	6	*CDKAL1*	20929641	C	0.37	0.30	1.38	0.01	0.42	0.33	1.49	0.15	0.28	0.26	1.08	0.79	0.36	0.30	1.34	1.08	1.65	0.008	0.33	0.01
rs11466493	3	*TGFBR2*	30661786	G	0.01	0.04	0.14	0.002	0.06	0.05	1.23	0.72	0.04	0.07	0.59	0.41	0.02	0.05	0.41	0.21	0.80	0.009	0.35	0.01
rs17025862	3	*TGFBR2*	30660132	G	0.01	0.04	0.14	0.002	0.06	0.05	1.23	0.72	0.04	0.07	0.59	0.41	0.02	0.05	0.41	0.21	0.80	0.009	0.35	0.01
rs7747989	6	*CDKAL1*	20921354	C	0.37	0.30	1.35	0.02	0.44	0.33	1.60	0.09	0.28	0.26	1.08	0.79	0.36	0.30	1.33	1.07	1.65	0.009	0.36	0.01
rs2328572	6	*CDKAL1*	21323815	A	0.003	0.02	0.15	0.03	0	0.03	NA	NA	0.01	0.03	0.48	0.48	0.005	0.02	0.19	0.05	0.69	0.01	0.44	0.01
rs10214694	6	*CDKAL1*	21321533	T	0.003	0.02	0.15	0.03	0	0.03	NA	NA	0.01	0.03	0.48	0.48	0.005	0.02	0.19	0.05	0.69	0.01	0.45	0.01
rs12190651	6	*CDKAL1*	21020651	G	0.05	0.03	1.38	0.30	0.12	0.02	6.17	0.0001	0.01	0.01	1.29	0.83	0.05	0.03	1.85	1.14	3.01	0.01	0.51	0.02
rs2275729	10	*HHEX*	94442410	G	0.15	0.12	1.32	0.13	0.15	0.12	1.29	0.51	0.21	0.10	2.26	0.02	0.16	0.12	1.44	1.08	1.92	0.01	0.52	0.02
rs1511024	4	*FABP2*	120459629	T	0.05	0.03	1.74	0.07	0.05	0.02	3.00	0.12	0.06	0.03	1.76	0.35	0.05	0.03	1.86	1.13	3.06	0.01	0.55	0.01
rs9350294	6	*CDKAL1*	20978072	T	0.38	0.32	1.26	0.07	0.44	0.32	1.65	0.07	0.33	0.29	1.24	0.45	0.38	0.32	1.31	1.05	1.61	0.01	0.55	0.01

The minor allele is the effect allele for the ORs. Chr, chromosome; L95, lower 95% CI boundary; U95, upper 95% CI boundary; NA, not available.

TABLE 3. Top Association Results in European-American Samples with Cases Defined as ETDRS Grade ≥30 for Variants in Genes Previously Associated with DR, DN, or T2D

					ARIC				CHS				MESA				CMH Combined Analysis					P Correction	
					MAF				MAF				MAF				MAF						
SNP	Chr	Gene	BP	Minor Allele	Cases	Controls	OR	P	Cases	Controls	OR	P	Cases	Controls	OR	P	Cases	Controls	OR	L95	U95	P	Bonferroni	Permutation
rs6133	1	*SELP*	167831970	A	0.04	0.12	0.34	0.002	0	0.12	NA	NA	0.09	0.12	0.74	0.69	0.05	0.12	0.32	0.17	0.58	2.3×10^{-4}	0.009	0.0002
rs3917779	1	*SELP*	167837472	A	0.04	0.12	0.35	0.003	0	0.12	NA	NA	0.09	0.12	0.74	0.69	0.05	0.12	0.32	0.17	0.59	2.9×10^{-4}	0.01	0.0001
rs9926180	16	*FTO*	52486108	T	0.33	0.25	1.51	0.01	0.38	0.25	1.81	0.08	0.32	0.17	2.32	0.07	0.34	0.24	1.65	1.24	2.18	5.0×10^{-4}	0.02	0.001
rs7500562	16	*FTO*	52488391	C	0.33	0.25	1.51	0.01	0.38	0.25	1.81	0.08	0.32	0.17	2.27	0.08	0.34	0.24	1.64	1.24	2.17	5.3×10^{-4}	0.02	0.001
rs12149433	16	*FTO*	52485580	G	0.15	0.09	1.75	0.01	0.20	0.08	2.84	0.01	0.05	0.04	1.07	0.95	0.15	0.09	1.93	1.33	2.81	5.5×10^{-4}	0.02	0.001
rs6128	1	*SELP*	167829528	T	0.08	0.16	0.49	0.009	0.05	0.17	0.26	0.05	0.09	0.18	0.45	0.27	0.08	0.16	0.44	0.27	0.70	5.6×10^{-4}	0.02	0.001
rs13335343	16	*FTO*	52488941	A	0.15	0.09	1.74	0.01	0.20	0.08	2.84	0.01	0.05	0.04	1.07	0.95	0.15	0.09	1.92	1.32	2.80	6.1×10^{-4}	0.02	0.001
rs12935710	16	*FTO*	52500306	T	0.32	0.23	1.55	0.01	0.33	0.25	1.48	0.28	0.27	0.15	2.09	0.14	0.32	0.22	1.61	1.21	2.14	0.001	0.04	0.002
rs10852525	16	*FTO*	52496582	A	0.17	0.12	1.52	0.05	0.20	0.10	2.29	0.05	0.09	0.05	2.09	0.34	0.17	0.11	1.71	1.20	2.45	0.003	0.12	0.005
rs9939152	16	*FTO*	52496904	G	0.33	0.25	1.45	0.03	0.35	0.26	1.53	0.23	0.32	0.20	1.89	0.18	0.33	0.25	1.52	1.15	2.02	0.004	0.14	0.006
rs4783824	16	*FTO*	52509162	T	0.13	0.09	1.50	0.09	0.20	0.08	2.95	0.01	0.05	0.05	1.69	0.62	0.14	0.08	1.78	1.21	2.64	0.004	0.15	0.005
rs1362570	16	*FTO*	52491048	C	0.14	0.10	1.52	0.07	0.20	0.09	2.45	0.03	0.05	0.04	1.07	0.95	0.14	0.09	1.70	1.16	2.48	0.007	0.26	0.009
rs4611524	17	*FTO*	58945384	T	0.31	0.41	0.67	0.02	0.33	0.42	0.68	0.27	0.36	0.40	0.86	0.74	0.32	0.41	0.69	0.52	0.91	0.009	0.34	0.009
rs12708942	16	*FTO*	52503705	A	0.14	0.10	1.46	0.1	0.20	0.09	2.63	0.02	0.05	0.04	1.15	0.89	0.14	0.09	1.68	1.14	2.47	0.009	0.35	0.01
rs9806929	16	*FTO*	52507417	A	0.14	0.10	1.46	0.1	0.20	0.09	2.63	0.02	0.05	0.04	1.15	0.89	0.14	0.09	1.68	1.14	2.47	0.009	0.35	0.01
rs551238	7	*EPO*	100159464	G	0.34	0.40	0.75	0.08	0.20	0.40	0.37	0.01	0.32	0.36	0.82	0.67	0.31	0.40	0.69	0.52	0.91	0.009	0.37	0.01

The minor allele is the effect allele for the ORs. Chr, chromosome; L95, lower 95% CI boundary; U95, upper 95% CI boundary; NA, not available.

TABLE 4. CHM Association Results for *SELP* SNPs in non-European-American CARe Populations, with Cases Defined as ETDRS Grade ≥14

SNP	Minor Allele	MAF Cases	MAF Controls	OR	L95	U95	P
African American							
rs6128	T	0.5	0.46	1.17	0.95	1.44	0.14
rs6133	C	0.42	0.45	0.89	0.72	1.09	0.26
rs3917779	G	0.48	0.50	0.94	0.76	1.15	0.55
Hispanic American							
rs6128	T	0.29	0.26	1.14	0.74	1.75	0.55
rs6133	A	0.16	0.18	0.91	0.55	1.52	0.72
rs3917779	A	0.14	0.16	0.89	0.52	1.52	0.67
Asian American							
rs6128	C	0.24	0.27	0.86	0.40	1.87	0.70
rs6133	A	0	0	NA	NA	NA	NA
rs3917779	A	0	0	NA	NA	NA	NA

L95, lower 95% CI boundary; U95, upper 95% CI boundary; NA, not available.

top association results for DR defined as ETDRS grade ≥14 and ETDRS grade ≥30 are shown in Tables 5 and 6, respectively. The lambdas for the quantile–quantile (Q–Q) plots were 1.01 and 1.00, respectively; Fig. 1). Supplementary Table S5 (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental) shows the results from the principal components analysis. There was no evidence of population stratification. Several variants in the two analyses had associations that were significant after Bonferroni correction. One variant, rs6856425, was significantly associated with DR in all three CARe cohorts for both definitions of DR with $P = 2.1 \times 10^{-5}$ after Bonferroni correction in the ETDRS grade ≥30 analysis. This association could not be replicated in the CARe African American cohorts (MAF 18%, OR = 0.94, $P = 0.69$).

We pursued replication of top findings from Tables 5 and 6 in independent cohorts of European ancestry with a fixed-effects meta-analysis model adjusted for age and sex (Table 7, Supplementary Fig. S1, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). None of the variants achieved significance in replication ($P < 1 \times 10^{-6}$). The smallest P value was for rs35260 ($P = 0.03$). For all SNPs examined in replication, there was a significant amount of heterogeneity ($P < 0.05$ for Q test). We performed a sensitivity analysis by removing the FinnDiane and Go-DARTS cohorts—the former because it had type 1 diabetes participants exclusively and both because they did not use ETDRS grading consistently for phenotyping (Supplementary Table S6, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). However, none of the associations were statistically significant in this analysis either, and significant heterogeneity remained for all SNPs with the exceptions of rs3917779 (Q test P = 0.06) and rs6856425 for the ETDRS grade ≥14 analysis (Q test P = 0.26). Given the significant residual heterogeneity, we then used a random effects model but found no significant difference in the results. We also performed meta-analyses without the CARe cohorts; there was no statistically significant result or any significant heterogeneity in these analyses. Of note, meta-analyses of the CARe cohorts alone also showed no significant heterogeneity.

In addition to the above replication efforts in European cohorts, we pursued replication of the same findings in two Asian cohorts, SiMES and SP2. None of the SNPs was statistically significant in these populations. We also investigated the *FTO* association in Go-DARTS; neither rs9926180 nor rs12935710 was significantly associated ($P = 0.84$ and 0.23, respectively).

Table 5. Top Association Results in European-American Samples with Cases Defined as ETDRS Score ≥14 for Variants in Genes on the IBC Chip

					ARIC				CHS				MESA				CMH Combined Analysis						*P* correction	
					MAF				MAF				MAF				MAF							
SNP	Chr	Gene	Position	Minor Allele	Cases	Controls	OR	*P*	Cases	Controls	OR	*P*	Cases	Controls	OR	*P*	Cases	Controls	OR	L95	U95	*P*	Bonferroni	Permutation
rs9332570	1	*F5*	167794699	G	0.09	0.18	0.44	8.8×10^{-5}	0.08	0.19	0.35	0.02	0.11	0.21	0.46	0.05	0.09	0.19	0.43	0.31	0.61	1.1×10^{-6}	4.3×10^{-5}	0.0001
rs7545236	1	*F5*	167796694	C	0.09	0.18	0.44	9.4×10^{-5}	0.08	0.19	0.35	0.02	0.11	0.21	0.46	0.05	0.09	0.19	0.44	0.31	0.61	1.2×10^{-6}	4.6×10^{-5}	0.0001
rs2298908	1	*F5*	167805168	T	0.09	0.18	0.45	0.0001	0.08	0.19	0.35	0.02	0.11	0.21	0.46	0.05	0.09	0.19	0.44	0.31	0.61	1.3×10^{-6}	4.9×10^{-5}	0.0001
rs1894701	1	*F5*	167797210	C	0.09	0.18	0.44	8.8×10^{-5}	0.08	0.19	0.36	0.03	0.11	0.21	0.46	0.05	0.09	0.19	0.44	0.31	0.61	1.3×10^{-6}	5.0×10^{-5}	0.0001
rs6427203	1	*F5*	167795655	C	0.09	0.18	0.45	0.0001	0.08	0.19	0.36	0.03	0.11	0.21	0.46	0.05	0.09	0.18	0.44	0.32	0.62	1.8×10^{-6}	7.1×10^{-5}	0.0001
rs35260	5	*PDE4D*	58668304	A	0.38	0.47	0.68	0.0003	0.36	0.48	0.61	0.07	0.24	0.50	0.31	7.4×10^{-5}	0.35	0.48	0.60	0.48	0.74	2.0×10^{-6}	7.6×10^{-5}	0.0001
rs6128	1	*SELP*	167829528	T	0.08	0.16	0.48	0.0007	0.06	0.18	0.30	0.02	0.08	0.20	0.36	0.02	0.08	0.17	0.43	0.31	0.62	3.1×10^{-6}	0.0001	0.0001
rs7168655	15	*PCSK6*	99782444	A	0.43	0.32	1.60	0.0002	0.47	0.33	1.76	0.04	0.43	0.33	1.50	0.13	0.44	0.33	1.61	1.31	1.98	7.6×10^{-6}	0.0003	0.0001
rs35259	5	*PDE4D*	58669348	T	0.39	0.47	0.70	0.005	0.36	0.48	0.61	0.07	0.26	0.50	0.36	0.0004	0.36	0.48	0.62	0.50	0.77	8.6×10^{-6}	0.0003	0.0001
rs6133	1	*SELP*	167831970	A	0.05	0.12	0.39	0.0003	0.03	0.13	0.22	0.02	0.07	0.13	0.50	0.16	0.05	0.12	0.38	0.25	0.59	1.1×10^{-5}	0.0004	0.0001
rs9332579	1	*F5*	167791936	T	0.06	0.14	0.37	8.9×10^{-5}	0.05	0.13	0.32	0.05	0.10	0.14	0.66	0.34	0.06	0.14	0.41	0.28	0.61	1.3×10^{-5}	0.0005	0.0001
rs3917779	1	*SELP*	167837472	A	0.05	0.12	0.40	0.0004	0.03	0.12	0.23	0.03	0.07	0.13	0.50	0.16	0.05	0.12	0.39	0.26	0.60	1.6×10^{-5}	0.0006	0.0002
rs6856425	4	*Near IDUA*	966918	C	0.04	0.02	2.36	0.01	0.11	0.02	6.17	0.0004	0.06	0.03	2.28	0.19	0.05	0.02	2.88	1.75	4.73	2.9×10^{-5}	0.001	0.0002
rs1824159	5	*PDE4D*	58654401	T	0.15	0.09	1.87	0.0006	0.11	0.09	1.15	0.76	0.19	0.08	2.81	0.004	0.15	0.09	1.88	1.39	2.53	3.5×10^{-5}	0.001	0.0001
rs1006273	15	*PCSK6*	99782862	A	0.52	0.43	1.47	0.002	0.59	0.43	1.93	0.02	0.51	0.42	1.48	0.14	0.53	0.43	1.53	1.24	1.88	5.6×10^{-5}	0.002	0.0001
rs16879334	5	*MTRR*	7944506	G	NA	NA	NA	NA	0.11	0.03	3.33	0.01	0.10	0.02	5.88	0.001	0.10	0.03	4.11	2.05	8.21	6.4×10^{-5}	0.002	0.0003
rs10036406	5	*PDE4D*	58588070	C	0.16	0.10	1.82	0.0007	0.12	0.10	1.29	0.55	0.18	0.09	2.14	0.04	0.16	0.10	1.78	1.34	2.38	9.1×10^{-5}	0.004	0.0001
rs7105871	11	*CADM1*	114717935	C	0.14	0.26	0.46	6.6×10^{-6}	0.24	0.25	0.98	0.94	0.25	0.28	0.87	0.65	0.17	0.26	0.59	0.46	0.77	9.4×10^{-5}	0.004	0.0002

The minor allele is the effect allele for the ORs. Chr, chromosome; L95, lower 95% CI boundary; U95, upper 95% CI boundary; NA, not available.

Table 6. Top Association Results in European-American Samples with Cases Defined as ETDRS Grade ≥30 for Variants within All Genes on the IBC Chip

					ARIC				CHS				MESA				CMH Combined Analysis						*P* Correction	
					MAF				MAF				MAF				MAF							
SNP	Chr	Gene	Position	Minor Allele	Cases	Controls	OR	*P*	Cases	Controls	OR	*P*	Cases	Controls	OR	*P*	Cases	Controls	OR	L95	U95	*P*	Bonferroni	Permutation
rs6856425	4	Near *IDUA*	966918	C	0.05	0.02	3.01	0.002	0.15	0.02	8.50	1.8×10^{-5}	0.09	0.03	3.54	0.10	0.07	0.02	3.83	2.26	6.46	5.3×10^{-7}	2.1×10^{-5}	0.0001
rs1555179	6	*MTHFDIL*	151313049	A	0.19	0.35	0.45	3.0×10^{-5}	NA	NA	NA	NA	0.18	0.32	0.47	0.17	0.19	0.34	0.45	0.32	0.65	1.3×10^{-5}	0.0005	0.0002
rs7078169	10	*CUBN*	17202320	T	0.09	0.03	2.86	0.0002	0.10	0.03	3.71	0.02	0.05	0.05	0.93	0.94	0.09	0.04	2.66	1.65	4.27	5.3×10^{-5}	0.002	0.0001
rs3901540	5	*PDE4D*	58465490	T	0.15	0.28	0.44	0.0001	0.18	0.29	0.52	0.13	0.32	0.32	0.98	0.96	0.17	0.29	0.50	0.35	0.70	5.6×10^{-5}	0.002	0.0002
rs6663966	1	*GLIS1*	53842601	C	0.04	0.01	5.16	5.2×10^{-5}	0.00	0.01	NA	NA	0.09	0.02	6.46	0.01	0.04	0.01	4.15	2.07	8.34	6.2×10^{-5}	0.002	0.0009
rs4697843	4	*HS3ST1*	10704856	A	0.21	0.33	0.52	0.0005	0.20	0.33	0.51	0.10	0.18	0.31	0.50	0.21	0.20	0.33	0.52	0.38	0.72	6.3×10^{-5}	0.002	0.0001
rs4521758	8	*PRKDC*	48904123	T	0.17	0.11	1.59	0.03	0.30	0.11	3.47	0.0007	0.23	0.09	2.92	0.04	0.20	0.11	1.98	1.41	2.77	7.3×10^{-5}	0.003	0.0006
rs10752067	10	*CUBN*	17188898	C	0.09	0.04	2.76	0.0003	0.10	0.03	3.73	0.02	0.05	0.05	0.93	0.94	0.09	0.04	2.59	1.62	4.16	7.9×10^{-5}	0.003	0.0001
rs3917462	1	Near *SELE*	167958291	C	0.03	0.01	4.03	0.006	0.05	0.01	9.05	0.009	0.05	0.01	5.16	0.12	0.03	0.01	4.75	2.17	10.36	9.3×10^{-5}	0.004	0.001
rs2508450	11	*IL10RA*	117369039	T	0.34	0.44	0.65	0.008	0.28	0.49	0.40	0.01	0.23	0.44	0.38	0.06	0.32	0.45	0.58	0.44	0.76	0.0001	0.004	0.0001
rs2195611	2	*PDE1A*	182964812	C	0.59	0.44	1.83	0.0001	0.58	0.45	1.69	0.12	0.45	0.48	0.91	0.83	0.58	0.45	1.68	1.29	2.19	0.0001	0.005	0.0001
rs11163323	1	Near *PRKACB*	81759046	T	0.33	0.19	2.05	1.7×10^{-5}	0.23	0.23	0.98	0.96	0.23	0.19	1.24	0.69	0.30	0.20	1.76	1.32	2.35	0.0001	0.005	0.0002
rs11155762	6	*MTHFDIL*	151314489	T	0.19	0.35	0.45	3.0×10^{-5}	0.38	0.34	1.17	0.64	0.18	0.32	0.47	0.17	0.22	0.34	0.55	0.40	0.75	0.0001	0.006	0.0005
rs17186372	1	*NOTCH2*	120302474	G	NA	NA	NA	NA	0.25	0.07	4.28	0.0002	0.14	0.07	2.08	0.26	0.21	0.07	3.45	1.81	6.56	0.0002	0.006	0.001
rs6583305	3	*UBXN7*	197603018	T	0.47	0.36	1.63	0.002	0.55	0.34	2.39	0.008	0.36	0.36	1.00	0.99	0.48	0.35	1.66	1.27	2.16	0.0002	0.007	0.0003
rs6764533	3	*UBXN7*	197572861	A	0.47	0.36	1.63	0.002	0.55	0.34	2.36	0.009	0.36	0.36	1.00	0.99	0.48	0.35	1.65	1.27	2.15	0.0002	0.007	0.0004
rs11545078	8	*GGH*	64101318	A	0.02	0.10	0.15	0.0003	0.08	0.11	0.64	0.47	0.05	0.11	0.37	0.33	0.03	0.10	0.26	0.13	0.53	0.0002	0.008	0.0002

The minor allele is the effect allele for the ORs; Chr, chromosome; L95, lower 95% CI boundary, U95, upper 95% CI boundary, NA, not available.

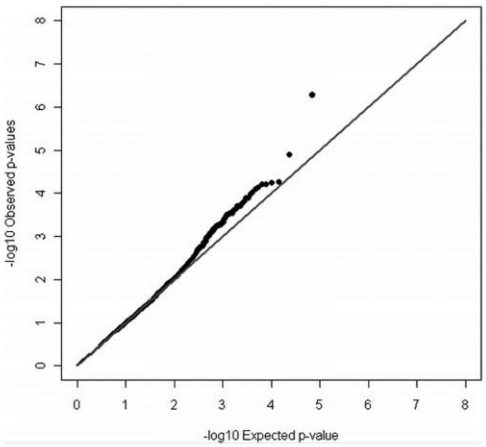

FIGURE 1. Quantile–quantile plot of all single SNPs examined on the IBC chip and their CMH association analysis to diabetic retinopathy, defined as an ETDRS score ≥30.

DISCUSSION

In this large international collaborative study, genes previously linked with T2D, DR, and DN and vascular diseases were not generally associated with DR. In the CARe European American population, among genes that have been previously associated with DR, DN, and T2D, three SNPs in *SELP* were associated with DR, even after adjustment for DR risk factors. However, we were unable to replicate this finding in other ethnic groups in CARe or in independent Caucasian cohorts. The *SELP* SNPs associated with DR in the present study were not in LD with rs6131, the SNP initially associated with diabetic albuminuria (Supplementary Fig. S2, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental).[64] P-selectin plays a role in leukocyte adhesion to endothelium during inflammation, and thus there is a biological rationale for its role in both diabetic microalbuminuria and retinopathy.[65] With regards to *FTO*, the SNPs associated with DR in CARe were not in significant LD with rs9939609, the SNP associated with T2D (Supplementary Fig. S2, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental).[66]

Importantly, we were unable to confirm an association to most genes that have previously been associated with DR and were included on the IBC chip. We note that for several of these genes, the IBC chip did not include SNPs in LD to the previously associated variants because the selection of tag SNPs may not have densely covered those genes (Supplementary Table S1, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). For some genes, most notably *EPO*, we did have excellent proxies to the initially reported variants. The chip included two perfect proxies (rs551238 and rs1734907) for the *EPO* SNP originally associated to PDR, rs1617640.[25] When we defined cases as having an ETDRS grade ≥30, rs551238 had a significant effect consistent with that found in the previous study, where the minor allele is protective (OR = 0.69, $P = 0.009$); the P value remains significant after correction by permutation ($P = 0.01$) but not by the Bonferroni method ($P = 0.37$). It is possible that with a larger sample size, the association would withstand the Bonferroni correction. We were also unable to detect an association to DR in other genes previously associated with DN and T2D. For the

TABLE 7. Replication Results in European Samples

			ARIC		CHS		MESA		AGES		BMES		FIND-Eye		FinnDiane		Go-DARTS		Lublin				Meta-analysis (Fixed Effects)	
Controls, *n*			732		160		140		249		175		627		570		774		620					
Cases ETDRS ≥14, *n*			153		33		36		92		67		105		2009		923		576					
Cases ETDRS ≥30, *n*			91		20		11		37		26		NA		1399		923		138					
SNP	Minor Allele	Definition of Cases	OR	*P*	OR	*P*	OR	*P*	MAF	OR	*P*	MAF	OR	*P*	MAF	OR	*P*	MAF	OR	*P*	MAF	OR	*P*	*z* Score	OR	*P*
rs9332570	G	ETDRS ≥14	0.43	0.0001	0.33	0.02	0.49	0.06	NA	0.75	0.26	0.21	1.24	0.42	0.19	0.79	0.26	0.27	1.16	0.05	NA	NA	NA	NA	0.002	1	0.99
rs35260	A	ETDRS ≥14	0.68	0.002	0.59	0.08	0.29	0.0001	NA	1.01	0.98	0.49	1.23	0.32	0.48	0.9	0.45	0.5	0.96	0.59	0.44	0.98	0.77	NA	NA	NA	−2.24	0.91	0.03
rs6128	T	ETDRS ≥14	0.48	0.0007	0.27	0.02	0.39	0.03	0.09	1.27	0.32	0.21	1.27	0.37	0.17	0.73	0.14	0.24	1.16	0.07	NA	1.13	0.21	0.16	1.01	0.96	0.6	1.03	0.55
rs71688655	A	ETDRS ≥14	1.59	0.0004	1.78	0.04	1.48	0.15	NA	0.91	0.57	0.34	0.96	0.83	0.36	1.08	0.63	0.34	1.02	0.82	0.34	0.96	0.6	NA	NA	NA	1.55	1.07	0.12
rs6133	A	ETDRS ≥14	0.59	0.0005	0.19	0.03	0.51	0.16	0.15	0.73	0.32	0.14	1.37	0.28	0.13	0.72	0.17	0.09	1.13	0.31	NA	1.08	0.45	0.08	1.03	0.84	−0.42	0.98	0.68
rs3917779	A	ETDRS ≥14	0.40	0.0007	0.20	0.03	0.51	0.16	0.09	0.73	0.32	0.14	1.25	0.45	0.12	0.72	0.19	0.09	1.14	0.29	NA	NA	NA	NA	NA	NA	−1.39	0.88	0.17
rs6856425	C	ETDRS ≥14	2.48	0.01	5.21	0.004	2.63	0.16	0.04	0.89	0.86	0.02	1.8	0.45	0.02	1.18	0.75	0.03	0.97	0.87	0.02	1.34	0.27	0.05	0.7	0.07	1.1	1.13	0.27
rs71058871	C	ETDRS ≥14	0.47	1.7×10^{-5}	0.97	0.91	0.86	0.62	NA	0.81	0.2	0.25	1.16	0.54	0.22	0.88	0.47	0.31	1	0.96	NA	NA	NA	NA	NA	NA	−1.85	0.9	0.07
rs6856425	C	ETDRS ≥30	3.19	0.003	6.8	0.002	3.25	0.19	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.03	0.79	0.39	0.02	1.34	0.27	0.05	0.46	0.06	1.65	1.28	0.1

All results are adjusted for age and sex. NA, not available.

DN genes, again the IBC chip may not have included SNPs in LD with the previously reported variants. For T2D, however, the IBC chip variants were specifically those previously associated at genome-wide significant levels.

Another explanation for the inability to replicate previous DR associations lies in the heterogeneity among studies regarding DR definitions and participants' mean duration of diabetes. We attempted to mitigate the heterogeneity of DR definitions by examining two different definitions. However, this may not be sufficient to account for all the possible phenotype heterogeneity. The studies from which we selected genes deemed to be previously associated with DR all used controls that were diabetic patients without DR, as we did; however, some of them were performed in type 1 diabetic patients, which is another potential source of heterogeneity. There was also great variability in the duration of T2D among cases and controls in CARe cohorts. In particular, there were participants with short durations of diabetes who were included. There is the potential for misclassification of controls if these participants did not have DR at the time of study inclusion but are at risk for significant DR with longer duration of diabetes. We attempted to correct for this by including duration of disease as a covariate in logistic regression, but these issues could still bias the results toward the null. However, our ability to detect an association with *EPO* indicates that the amount of control misclassification in CARe is not significant enough to prevent detection of associations of this effect size.[25,67] Finally, while the current investigation is the largest candidate gene study for DR to date, it still has limited power to detect genetic associations of modest or small effects (Supplementary Table S7, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). In particular, CARe includes a modest number of European-American cases (122) defined as ETDRS grade ≥30. Examining milder degrees of DR as outcomes may have further decreased our ability to detect associations as the development of early DR has a lower heritability. Of note, the ARIC cohort is larger than the other cohorts, and the top findings in the analyses were often driven by the results in ARIC.

In the second phase of the analysis, we took an unbiased approach at the remaining genes available on the IBC chip. Although we found several strong associations in our discovery cohort, replication in independent samples did not yield variants with consistent effects nor any variants that met the replication significance threshold ($P = 1 \times 10^{-6}$). The rs6856425 association was initially compelling because it was consistent within each CARe cohort. Furthermore, the strength of the association was greater when DR was defined as ETDRS grade ≥30 vs. ETDRS grade ≥14, which is in line with the expected greater heritability of more advanced DR phenotypes. The rs35260 variant had the lowest P value in the replication meta-analyses ($P = 0.03$), but this was still far below the replication threshold for significance ($P = 1 \times 10^{-6}$).

Failure to replicate a genetic association can be explained broadly, either as a false positive in the discovery cohort or a false negative in the replication cohort. For rs6856425, a rare variant, the initial estimate was based on limited instances of the minor allele: 9 in cases and 20 in controls. Small numbers of observations can lead to unstable effect estimates that represent chance statistical fluctuations rather than true associations. This underscores the importance of large sample sizes in both discovery and replication cohorts, particularly for rare variants. Another possible reason for false positives is population stratification, but there was no significant population stratification in this study.

False negatives in the replication cohort can be due to a lack of power, genotyping/imputation imprecision or heterogeneity between cohorts. Our aggregate replication cohort sample was well powered (Supplementary Table S7, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-7510/-/DCSupplemental). Of note, imputation was used by most replication cohorts for at least one of the SNPs. Imputation quality scores were greater than 0.90 for all SNPs with the exception of rs6856425 in FIND-Eye where the imputation quality score was 0.67. Errors in imputed genotype calls could lead to false-negative results, particularly for rare SNPs, which are more susceptible to genotyping artifacts. In addition, there was significant heterogeneity among samples that could not be explained by excluding type 1 diabetes participants and cohorts that did not use ETDRS grading consistently. Because the heterogeneity was not present when meta-analysis was restricted to the replication cohorts alone or the CARe cohorts alone, the "winner's curse" effect of large effect sizes in the discovery cohorts likely explains most of this heterogeneity.[68] Some heterogeneity might also derive from the different DR ascertainment methods and case–control definitions. Cohorts differed in their photography protocols, with some cohorts having one field of only one eye for phenotype determination. This introduces misclassification bias for participants for whom the DR grade in the one or two fields photographed may not accurately represent the DR grade in other fields or the contralateral eye. It is therefore possible that some of the variants associated with DR in CARe may eventually be replicated in larger studies with direct genotyping and better phenotype harmonization.

In summary, in this candidate gene analysis of DR with data from the CARe consortium, with replications in a several large, well-powered samples, we found little evidence of a major DR gene. This is the largest number of candidate genes studied for DR to date. Although no association could be confirmed with a high threshold for significance, the results are hypothesis generating and the genes associated with DR in CARe could be prioritized in studies. The importance of well-powered replication and phenotype harmonization are highlighted by our study. These issues will continue to be important as results from genome-wide association studies for DR become increasingly available.

Acknowledgments

The authors thank Joan W. Miller and David M. Altshuler for their support of the study.

References

1. National Institute of Diabetes and Digestive and Kidney Diseases. National diabetes statistics fact sheet: general information and national estimates on diabetes in the United States. 2000. Bethesda, MD: U.S. Department of Health and Human Services, National Institute of Health, Publication No. 02-3892.
2. Klein R, Klein BE, Moss SE, Cruickshanks KJ. The Wisconsin Epidemiologic Study of Diabetic Retinopathy, XVII: The 14-year incidence and progression of diabetic retinopathy and associated risk factors in type 1 diabetes. *Ophthalmology.* 1998;105:1801–1815.
3. Klein BE. Overview of epidemiologic studies of diabetic retinopathy. *Ophthalmic Epidemiol.* 2007;14:179–183.
4. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet.* 2010;376:124–136.
5. Zhang X, Saaddine JB, Chou CF, et al. Prevalence of diabetic retinopathy in the United States, 2005–2008. *JAMA.* 2010;304:649–656.
6. Nathan DM. Long-term complications of diabetes mellitus. *N Engl J Med.* 1993;328:1676–1685.
7. Klein R, Klein BE, Moss SE, Cruickshanks KJ. Relationship of hyperglycemia to the long-term incidence and progression of diabetic retinopathy. *Arch Intern Med.* 1994;154:2169–2178.

8. Looker HC, Nelson RG, Chew E, et al. Genome-wide linkage analyses to identify loci for diabetic retinopathy. *Diabetes.* 2007; 56:1160–1166.
9. Hietala K, Forsblom C, Summanen P, Groop PH. Heritability of proliferative diabetic retinopathy. *Diabetes.* 2008;57:2176–2180.
10. Arar NH, Freedman BI, Adler SG, et al. Heritability of the severity of diabetic retinopathy: the FIND-Eye study. *Invest Ophthalmol Vis Sci.* 2008;49:3839–3845.
11. Abhary S, Hewitt AW, Burdon KP, Craig JE. A systematic meta-analysis of genetic association studies for diabetic retinopathy. *Diabetes.* 2009;58:2137–2147.
12. Liew G, Klein R, Wong TY. The role of genetics in susceptibility to diabetic retinopathy. *Int Ophthalmol Clin.* 2009;49:35–52.
13. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet.* 2003;33:177–182.
14. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med.* 2002;4: 45–61.
15. Kottgen A, Hwang SJ, Rampersaud E, et al. TCF7L2 variants associate with CKD progression and renal function in population-based cohorts. *J Am Soc Nephrol.* 2008;19:1989–1999.
16. Krentz AJ, Clough G, Byrne CD. Interactions between microvascular and macrovascular disease in diabetes: pathophysiology and therapeutic implications. *Diabetes Obes Metab.* 2007;9:781–791.
17. Ohno T, Kinoshita O, Fujita H, et al. Detecting occult coronary artery disease followed by early coronary artery bypass surgery in patients with diabetic retinopathy: report from a diabetic retino-coronary clinic. *J Thorac Cardiovasc Surg.* 2010;139:92–97.
18. Kofler B, Mueller EE, Eder W, et al. Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study. *BMC Med Genet.* 2009;10:35.
19. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med.* 2008;359:1577–1589.
20. Cheung N, Wong TY. Diabetic retinopathy and systemic vascular complications. *Prog Retin Eye Res.* 2008;27:161–176.
21. Kawasaki R, Cheung N, Islam FM, et al. Is diabetic retinopathy related to subclinical cardiovascular disease? *Ophthalmology.* 2011;118:860–865.
22. Musunuru K, Lettre G, Young T, et al. Candidate gene association resource (CARe): design, methods, and proof of concept. *Circ Cardiovasc Genet.* 2010;3:267–275.
23. Keating BJ, Tischfield S, Murray SS, et al. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One.* 2008;3:e3583.
24. Uhlmann K, Kovacs P, Boettcher Y, Hammes HP, Paschke R. Genetics of diabetic retinopathy. *Exp Clin Endocrinol Diabetes.* 2006;114:275–294.
25. Tong Z, Yang Z, Patel S, et al. Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. *Proc Natl Acad Sci U S A.* 2008;105:6998–7003.
26. Ng DP, Tai BC, Lim XL. Is the presence of retinopathy of practical value in defining cases of diabetic nephropathy in genetic association studies?—the experience with the ACE insertion/deletion polymorphism in 53 studies comprising 17,791 subjects. *Diabetes.* 2008;57:2541–2546.
27. Freedman BI, Bostrom M, Daeihagh P, Bowden DW. Genetic factors in diabetic nephropathy. *Clin J Am Soc Nephrol.* 2007;2: 1306–1316.
28. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet.* 2007;8:657–662.
29. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* 2007; 445:881–885.
30. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science.* 2007;316:1331–1336.
31. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007;316:889–894.
32. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40:638–645.
33. Sandhu MS, Weedon MN, Fawcett KA, et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet.* 2007;39:951–953.
34. Winckler W, Weedon MN, Graham RR, et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes.* 2007;56:685–693.
35. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol.* 2002;156: 871–881.
36. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989;129: 687–702.
37. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1:263–276.
38. Wilson JG, Rotimi CN, Ekunwe L, et al. Study design for genetic analysis in the Jackson Heart Study. *Ethn Dis.* 2005;15:S6–30-37.
39. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 2003;26(suppl 1):S5–S20.
40. Wong TY, Klein R, Islam FM, et al. Diabetic retinopathy in a multi-ethnic cohort in the United States. *Am J Ophthalmol.* 2006; 141:446–455.
41. Cheung N, Wang JJ, Rogers SL, et al. Diabetic retinopathy and risk of heart failure. *J Am Coll Cardiol.* 2008;51:1573–1578.
42. Klein R, Marino EK, Kuller LH, et al. The relation of atherosclerotic cardiovascular disease to retinopathy in people with diabetes in the Cardiovascular Health Study. *Br J Ophthalmol.* 2002;86:84–90.
43. Diabetic retinopathy study. Report Number 6. Design, methods, and baseline results. Report Number 7. A modification of the Airlie House classification of diabetic retinopathy. Prepared by the Diabetic Retinopathy. *Invest Ophthalmol Vis Sci* 1981;21:1–226.
44. Ojaimi E, Nguyen TT, Klein R, et al. Retinopathy signs in people without diabetes the multi-ethnic study of atherosclerosis. *Ophthalmology.* 2011;118:656–662.
45. Colagiuri S, Lee CM, Wong TY, Balkau B, Shaw JE, Borch-Johnsen K. Glycemic thresholds for diabetes-specific retinopathy: implications for diagnostic criteria for diabetes. *Diabetes Care.* 2011;34: 145–150.
46. Wong TY, Liew G, Tapp RJ, et al. Relation between fasting glucose and retinopathy for diagnosis of diabetes: three population-based cross-sectional studies. *Lancet.* 2008;371:736–743.
47. Guillausseau PJ, Massin P, Charles MA, et al. Glycaemic control and development of retinopathy in type 2 diabetes mellitus: a longitudinal study. *Diabet Med.* 1998;15:151–155.
48. Stratton IM, Kohner EM, Aldington SJ, et al. UKPDS 50: risk factors for incidence and progression of retinopathy in type II diabetes over 6 years from diagnosis. *Diabetologia.* 2001;44:156–163.
49. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. *BMJ.* 1998;317:703–713.
50. Ferris FL 3rd, Chew EY, Hoogwerf BJ. Serum lipids and diabetic retinopathy. Early Treatment Diabetic Retinopathy Study Research Group. *Diabetes Care.* 1996;19:1291–1293.
51. Cochran WG. Some methods for strengthening the common X2 tests. *Biometrics.* 1954;10:417–451.
52. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22: 719–748.
53. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med.* 2005;24:2911–2935.
54. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–575.
55. Qiu C, Cotch MF, Sigurdsson S, et al. Retinal and cerebral microvascular signs and diabetes: the age, gene/environment susceptibility-Reykjavik study. *Diabetes.* 2008;57:1645–1650.

56. Mitchell P, Smith W, Wang JJ, Attebo K. Prevalence of diabetic retinopathy in an older community. The Blue Mountains Eye Study. *Ophthalmology.* 1998;105:406–411.
57. Doney AS, Leese GP, Olson J, Morris AD, Palmer CN. The Y402H variant of complement factor H is associated with age-related macular degeneration but not with diabetic retinal disease in the Go-DARTS study. *Diabet Med.* 2009;26:460–465.
58. Wong TY, Cheung N, Tay WT, et al. Prevalence and risk factors for diabetic retinopathy: the Singapore Malay Eye Study. *Ophthalmology.* 2008;115:1869–1875.
59. Yim-Lui Cheung C, Wong TY, Lamoureux EL, et al. C-reactive protein and retinal microvascular caliber in a multiethnic Asian population. *Am J Epidemiol.* 2010;171:206–213.
60. Buraczynska M, Baranowicz-Gaszczyk I, Tarach J, Ksiazek A. Toll-like receptor 4 gene polymorphism and early onset of diabetic retinopathy in patients with type 2 diabetes. *Hum Immunol.* 2009;70:121–124.
61. Cochran WG. Problems arising in the analysis of a series of similar experiments. *J R Stat Soc.* 1937;4:102–118.
62. Cochran WG. The combination of estimates from different experiments. *Biometrics.* 1954;10:101–129.
63. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol.* 2008;32:381–385.
64. Liu Y, Burdon KP, Langefeld CD, et al. P-selectin gene haplotype associations with albuminuria in the Diabetes Heart Study. *Kidney Int.* 2005;68:741–746.
65. Tedder TF, Steeber DA, Chen A, Engel P. The selectins: vascular adhesion molecules. *FASEB J.* 1995;9:866–873.
66. Freathy RM, Timpson NJ, Lawlor DA, et al. Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes.* 2008;57:1419–1426.
67. Abhary S, Burdon KP, Casson RJ, Goggin M, Petrovsky NP, Craig JE. Association between erythropoietin gene polymorphisms and diabetic retinopathy. *Arch Ophthalmol.* 2010;128:102–106.
68. Kraft P. Curses—winner's and otherwise—in genetic epidemiology. *Epidemiology.* 19:649–651, 2008; discussion 57–58.